



UNIVERSITÀ DI NAPOLI L'ORIENTALE

ANNALI

*del Dipartimento di Studi
Letterari, Linguistici e Comparati
Sezione linguistica*

AION-L

N.S. 12

2023



UniorPress

ANNALI

*del Dipartimento di Studi
Letterari, Linguistici e Comparati
Sezione Linguistica*

AION-L

N.S. 12
2023



Direttore/Editor-in-chief: *Alberto Manco*

Comitato scientifico/Scientific committee: *Ignasi-Xavier Adiego Lajara, Françoise Bader, Annalisa Baicchi, Philip Baldi, Giuliano Bernini, Carlo Consani, Pierluigi Cuzzolin, Paolo Di Giovine, Norbert Dittmar, Laura Gavioli, Nicola Grandi, Marco Mancini, Andrea Moro, Vincenzo Orioles, Paolo Poccetti, Diego Poli, Ignazio Putzu, Giovanna Rocca, Velizar Sadovski, Domenico Silvestri, Francisco Villar*

Comitato di redazione/Editorial board: *Anna Cardinaletti, Mario Cardona, Valeria Caruso, Marina Castagneto, Francesca Chiusaroli, Anna De Meo, Lucia di Pace, Francesca Dragotto, Elena Favilla, Gloria Gagliardi, Franco Lorenzi, Sabrina Machetti, Alberto Manco, Antonietta Marra, Johanna Monti, Andrea Nuti, Rossella Pannain, Lorenzo Spreafico, Massimo Vai, Iride Valenti*

Segreteria di redazione/Editorial assistant: *Valeria Caruso*
e-mail: segreteriaion@unior.it

Annali-Sezione Linguistica, c/o *Alberto Manco, Università di Napoli L'Orientale, Dipartimento di Studi Letterari, Linguistici e Comparati, Palazzo Santa Maria Porta Coeli, Via Duomo 219, 80138 Napoli - albertomanco@unior.it*

ISSN 2281-6585

Registrazione presso il Tribunale di Napoli n. 2901 del 9-1-1980

web: <http://www.serena.unina.it/index.php/aionlin/index>



This work is licensed under a Creative Commons Attribution 4.0 International License

UniorPress, Via Nuova Marina 59 - 80133 Napoli

Per la redazione delle proposte i collaboratori sono invitati ad attenersi con cura alle "norme" disponibili nel sito della rivista.

Le proposte di pubblicazione inviate alla rivista vengono valutate da revisori anonimi. A tal fine una loro copia dev'essere priva di qualunque riferimento all'autore.



UNIVERSITÀ DI NAPOLI L'ORIENTALE

ANNALI

*del Dipartimento di Studi
Letterari, Linguistici e Comparati
Sezione Linguistica*

AION-L

N.S. 12

2023



UniorPress

PROPRIETÀ RISERVATA

INDICE

LETTERE APERTE, DISCUSSIONI, PROPOSTE

Problemi e prospettive di ricerca, convegni e tavole rotonde, notizie e suggerimenti

- VELIZAR SADOVSKI, *Word-formation of epithets and names in Indo-Iranian in their relationship with underlying syntax, ritual phraseology and poetical idiomatics* 11

ARTICOLI, NOTE, SAGGI

Analisi linguistiche di testi arcaici, riflessioni su aspetti e problemi linguistici del mondo antico, linee e momenti di preistoria e protostoria linguistica

- CRISTINA RESMINI, *Grammaticalisation paths of the Proto-Indo-European anaphoric pronoun *tód across the Germanic languages* 41

Ricerche e problemi linguistici di ambito teorico e applicato

- LUISA CORONA, ROSALBA NODARI, *“Finisco da scrivere con le parole più belle”. Lessico e strategie linguistiche dell’oralità in Lettere da una tarantata* 135
- FRANCESCA D’ANGELO, *Assessing translation quality: a survey of research into human translation, post-editing and machine translation* 179
- LUCIA DI PACE, ROSSELLA PANNAIN, *Strategie linguistiche nella campagna elettorale delle politiche 2022. Una focalizzazione sull’impiego dei deittici di persona* 201
- FRANCO LORENZI, *Parlare di niente. Appunti per un lessico dell’inesistente* 253
- ELENA PEPPONI, *Comunicazione istituzionale ampia e tecnologie per il Trattamento Automatico del Linguaggio: possibili applicazioni e sviluppi* 275
- NICOLA SANTONI, *In search for the motivation of markedness through metaphorical processes. Chains and matrices with a reasonable degree of universality* 325

PAOLA VERNILLO, *MIP-IT: Identificazione delle metafore linguistiche in italiano. Una proposta preliminare tra questioni teoriche e problemi pratici* 345

BIBLIOGRAFIE, RECENSIONI, RASSEGNE

GIULIANA GIUSTI, VINCENZO NICOLÒ DI CARO, DANIEL ROSS (eds.), *Pseudo-coordination and Multiple Agreement Constructions*, Amsterdam / Philadelphia, John Benjamins, 2022 385

ELISABETTA BONVINO, DIEGO CORTÉS VELASQUEZ, ANNA DE MEO, ELISA FIORENZA, *Agire in L2. Processi e strumenti nella linguistica educativa*, Milano, Hoepli, 2023 403

ANDREA MORO, *La razza e la lingua. Sei lezioni sul razzismo*, Milano, La nave di Teseo editore, 2019 407

LETTERE APERTE, DISCUSSIONI, PROPOSTE

*Problemi e prospettive di ricerca,
convegni e tavole rotonde,
notizie e suggerimenti*

Word-formation of epithets and names in Indo-Iranian in their relationship with underlying syntax, ritual phraseology and poetical idiomatics

Abstract

L'articolo presenta correlazioni sistematiche tra composti e derivativi nominali (da un lato) e strutture sintattiche e collocazioni fraseologiche (dall'altro) nei Veda e nell'Avesta e argomenta a favore del loro background indoiranico comune. Nella Sezione 1 si dimostra che un nome personale antico-iranico, attestato nella trasmissione elamita da Persepoli, deriva da un composto iir. **H₁ṛta-mi(H)azd^hHa-*, e spiega che questo composto è costituito da un membro anteriore iir. **H₁ṛta-* 'Rightness/Giustizia' e un termine posteriore **mi(H)azd^hHa-* 'offerta di] pasto (solido)', presente anche in formazioni come l'avestico recente (*)*Nərə-miiazda-* (discusso nella Sezione 3) e l'antico-iran. **Dāma-miiazda-* (Sezione 4). Lo stesso termine **mi(H)az-d^hHa-* è composto da due membri composizionali e viene interpretato come un *nomen actionis* desintagmatico, che risale in ultima analisi a una frase verbale fattitiva. Di gran pregio per delineare il suo sviluppo sono i sintagmi attestati nell'avestico (antico e recente). Infatti, nella Sezione 2 si identificano importanti collocazioni formulaiche in cui lessemi come iir. **H₁ṛta-* e **mi(H)az-d^hHa-* ricorrono insieme nella liturgia avestica e in paralleli tratti dal rituale vedico. Elementi di tali collazioni nelle Gāṓā – av. *aṣā-*, *nāmah-*, *myazda-*, ma anche *nar-* e *dāman-* / radice *DĀ* – corrispondono pienamente ai termini antico-indiani *ṛtá-*, *nāmas-*, *mīyédha-*, ma anche *nár-* e *dhāman-* / radice *D^hĀ*; sono attestati in ulteriori composti, ad es. nell'antico-iran. *(*H*)*ṛta-d^hāma-* (Sezione 5). Il valore aggiunto di queste costellazioni lessicali sta nella loro compresenza, sia in composti nominali, sia in strutture fraseologiche; le cinque sezioni dell'articolo illustrano il ruolo di tali collocazioni formulaiche per la ricostruzione del lessico e della formazione di parola, ma anche della poesia e delle pratiche rituali dell'indoiranico.

Parole chiave: lessico indoiranico; formazione di parola nell'indoeuropeo; composti e derivativi nominali; "light verb constructions"; rituale vedico; liturgia avestica; mazdeismo; brāhmanismo

The present article demonstrates systematic correspondences between nominal compounds and derivatives, on the one hand, and syntactic structures and phraseological collocations, on the other, in the Veda and the Avesta and argues in favour of their com-

* Velizar Sadovski, Austrian Academy of Sciences, velizar.sadovski@oeaw.ac.at.

mon Indo-Iranian background. Part 1 shows that an Old Iran. personal name attested in Elamite transmission from Persepolis comes from an Ilr. **H₁ṛta-mi(H)azd^hHa-* and points out that this compound consists of an anterior member Ilr. **H₁ṛta-* ‘Rightness’ and a posterior term **mi(H)azd^hHa-* ‘(solid) meal offering’, also occurring in formations like Young Avestan (**)Nərə-miiazda-* (discussed in Part 3) and Old Iran. **Dāma-mijazda-* (Part 4). The term **mi(H)az-d^hHa-* itself consists of two compositional members and is analyzed as a desyntagmatic *nomen actionis* that ultimately goes back to a factitive verbal phrase. Valuable for its interpretation are syntagmata attested in Old and Young Avestan: Part 2 discovers important formulaic collocations in which lexemes like Ilr. **H₁ṛta-* ‘Rightness’ and **mi(H)az-d^hHa-* occur together in the Avestan liturgy and in parallels from the Vedic ritual. Elements of such collocations in the Gāṇās – Av. *aša-, nōmah-, myazda-* but also *nar-* and *dāman-* / root *DĀ* – fully correspond to the Old Indic terms *ṛtá-, námas-, miyédha-* but also *nár-* and *dhāman-* / root *D^hĀ*; they are attested in further compounds, e.g. Old Iran. **(H)ṛta-d^hāma-* (Part 5). The added value of these lexical constellations is in their joint occurrence in nominal compounds and in phraseological structures; the five parts of the article display the role of such formulaic collocations for the reconstruction of the Indo-Iranian lexicon, word formation but also ritual poetry and practice.

Keywords: indo-iranian lexicon; Indo-European word formation; nominal compounds and derivatives; light verb constructions; Vedic ritual; Avestan liturgy; Mazdeism; Brāhmaṇism

0.1. A series of studies starting with two papers given at the Conference *Indo-European Word Formation – Inventory of Analysis* at Copenhagen in 2000 and at the Colloquium of the Indogermanische Gesellschaft at Paris in 2003 – published in SADOVSKI 2004, developed in SADOVSKI 2007, 2012ab, 2013, 2016, 2018ab, 2019 [2020] and 2020 –, have been concerned with displaying systematic correspondences between (compounded) formations (adjectival epithets or nouns, viz. appellatives and proper names) and elements of free syntax and phraseology in Indo-Iranian and in a larger, Indo-European perspective.

0.2. The epithets and proper names attested in the Veda and the Avesta, but also in the Old West Iranian (i.e. Old Persian and Median) of both direct and indirect (‘collateral’) transmission exhibit surprisingly numerous systematic formal-and-semantic matches between nominal formations and syntactic units – esp. formulae and collocations from sacred poetry and law: There is hardly a better example of two Indo-European language traditions with so many interconnections between word-formation products and phra-

seological junctures, esp. between (compounded) cultic epithets and epicicles referring to language of ritual poetry that virtually concern all spheres of spiritual life: theology, cosmology, ‘anthropology’.

0.3. The multilingual corpus of attestations of Old (West) Iranian lexemes in sources from the Old Persian Great Kingdom offers further numerous instances of (compounded) proper names and epithets that are interpretable as nominalizations of cultic phraseology attested in the oldest Iranian and Indic sacred texts. The reconstruction of the Old Iranian form of such names and the analysis of their word-formation and semantic context opens new fascinating perspectives to the reconstruction of Indo-Iranian, for what concerns the cultural-historical backgrounds of processes of name-giving in their relationship with the sacred poetry of the Veda and Avesta.

Simultaneously, this material is able to enlarge our knowledge about syntactic expressions, in particular phraseological formulae which are (sometimes but scarcely) attested in the extant ritual texts and whose stability and role in the daily cultic practice can now be corroborated by their role as syntactic structures underlying such elements of the lexicon as epithets and personal names that more than anything else represent spiritual ideas and systems of values of the respective communities in a crystalized, concentrated form.

1. The multilingual corpus of attestations of Old (West) Iranian lexemes in sources from the Old Persian Great Kingdom offers numerous instances of (compounded) proper names, epithets and epicicles interpretable as nominalizations of cultic phraseology that we can discover in the oldest Iranian and Indic sacred texts. The reconstruction of the Old Iranian form of such names and the analysis of their word-formation, semantic context and syntactic combinatorics of the lexemes that form the terms of the respective compounds open new fascinating perspectives to the reconstruction of Indo-Iranian, for what concerns the cultural-historical backgrounds of processes of name-giving in their relationship with the sacred poetry of the Veda and Avesta.

The repertory of possibly Iranian names in Elamite transmission contains the male name *Irdamiyašda* <*Ir-da-mi-ya-iš-da*>. It is attested as a hapax, on the Persepolis fortification tablet 975:2 (cf. HALLOCK 1969: 703a).

1.0. Following BENVENISTE 1966: 84 (“**ṛta-myazda*- avec av. *myazda*- « banquet rituel »”), MAYRHOFER 1973: 164, #8.590, supports the interpre-

tation of this string as Iran. **Ṛta-miāzda-* (with refs. to obsolete interpretations); both of them refrain from indicating a reconstructed meaning of the formation. While TAVERNIER 2007 does not mention compounds with **miāzda-*, ZADOK 2009: 153, #233, discusses <*Ir-da-mi-ya-iš-da*> as a possible but not certain reconstruct of the second term in <[D]a-²ma-¹mi-as/z-ta>, <[D]a-²ma-mi_j-as/z-ta> (on which s. § 4.).

1.1. On formal grounds, OIran. *(H)*ṛta-miāzda-* < IIr. **Hṛta-mi(H)azd^hHa-* appears to be the most economic analysis of the attested formation (in accord with the phonological correspondences between Elamite and Old Iranian established in MAYRHOFER 1973: 15ff.). Still, this reconstruction is to be tested on criteria such as (A.) presumable meaning of the entire formation, (B.) independent occurrence of the first and of the second term of this compound as anterior/posterior members of other compounded formations, and (C.) possible combinations of the presumed compositional terms as free lexemes in syntactic and phraseological contexts external to the compound itself:

1.2. A compound proper name **Hṛta-mi(H)azd^hHa-* is interpretable, for both onomastic and compound-typological reasons,¹ as ‘(one) who has/makes (solid) meal offerings to Rightness’, a stative/factitive *bahuvrihi* (and not a *tatpuruṣa* ‘meal of Rightness’), with first term IIr. **Hṛtá-* ‘rightness’, Ved. *ṛtá-* ‘id.’. In (Y)Av. the (non-stressed) zero-grade form °*arata-* is attested as second term in compounded personal names such as *uxšiiat.arata-* ‘the one who lets R. grow’, *astuuat.arata-* ‘may Rightness be osseous/material[ized]’ etc.; in both Avestan dialects we have “*guṇa*” formations such as (not root-stressed) *dājīr.arata-* ‘oppressor of Rightness’, YAv. (Y. 12,4), *an-arata-* ‘having no Rightness, adversary of R.’, °*arata-* < *(H)*arta-* in GAv. (Y. 53,6; 53,9) and, otherwise, (root-stressed) G/YAv. *aša-*, n. ‘r/Rightness’ < *(H)*árta-*. The meaning ‘rightness’ includes both ‘rightness, regularity, (cosmic) order’ and ‘rightness, correctness, truth’ (recently cf. SADOVSKI 2017, with refs.).

1.3. While the anterior member of the compound seems clear, the second term **miāzda-* ‘(solid) ritual meal, sacrificial repast’ is relatively rarely attested in compounds. OIran. **mi(H)azda-*, Ved. *miyédha-* go back to an IIr. **mi(H)*

¹ I thank Rüdiger SCHMITT and Alexander LUBOTSKY for kindly reading an earlier draft of this paper.

as-d^hHa-, a compound of the type Ilr. **X-d^hH-a-* < IE **X-d^hh₁-o-* ‘characterized by setting/doing X’, most recently presented in SADOVSKI 2018b, with refs.²

1.3.1. There are several possible scenarios regarding the etymology of the first element, **mijās*°. (a.) One of them operates with YAv. *maiiāh-* in a presumed meaning ‘lust, (physical) enjoyment’ and reconstructs an Ilr. **máiHas-* (LUBOTSKY forthc., s.v.) < **méjHes-*, connecting it to Lat. *mī-tis* ‘mellow, soft’, OIran. *mín* ‘smooth, soft’, Lith. *mīe-las* ‘dear, lovely, tender’ etc., with additional suffixal elements. In this version, **mi-i-as-* would go back to a R(z) ablaut variant **mi(H)as-*. (b.) A sub-variant of this approach takes **méj(H)es-* to be the basis of the super-zero-grade compound **mi(H)s-d^hh₁-ó-* ‘reward’, interpreting it as an orig. ‘setting/making a joy for sb.’ (so HINTZE 2000: 67; LUBOTSKY, *ib.*, s.v. **mižd^h(H)a-*, prefers etymologizing IE **mis-* from “**mei-* ‘to exchange’ [...], if the root does not contain an initial **h₂*, rather than **mei(H)-es-*”); I briefly discuss pros and contras of a “vet” reconstruct Ilr. **máġ(H)as-* in SADOVSKI forthc. and will skip this argument here. A further reconstruction approach may start from an *-s*-stem Ilr. **máġas-* < an IE transponate **méġ-es-* going back to an aniṭ root, **meġ*. Traditionally, we distinguish two such roots in Ilr.: **1maġ/mi* ‘to strengthen/support/fortify’ (SADOVSKI 2016: 731f.; STEER, *HS* 120, 144: *(H)*meġ-*), Ved. *mi-*, *minoti*, and **2maġ/mi* ‘to exchange; to alter(nate)’. The second meaning could connect Ilr. **mijās-d^hHa-* ‘(ritual) meal [offering]’ and **mis-d^hH-á-* ‘reward’ by a common meaning of *‘giving in exchange’: problematic remains, however, that the reconstruction of **2maġ/mi* mostly operates with an initial laryngeal, as Ilr. **Hmaġ* < IE **h₂meġ*, because of Gr. *ἀμείβω* (< **h₂mei-g^{wo}*); but **h₂-* is absent in Gk. *μισθός*. If Ilr. **mi(i)as-d^hHa-* ‘(ritual) meal [offering]’ (theoretically even **mis-d^hH-á-* ‘reward’ but cf. SCHUTZEICHEL 2013: 130, SADOVSKI forthc., with refs.) could be connected with Ved. *máyas-*, n. ‘strengthening, support(ing), refreshment, meal; (physical) enjoyment’, these meanings are also presumable for YAv. *maiiāh-* (orig. from **1maġ/mi* ‘to strengthen, to fortify’?). If we part from the premise that various roots Ilr. **maġ*, *[H]*maġ* and **maġ(H)* could have become object of contamination

² A n. *actionis/abstr.* in Ilr. **-d^hH-i-* ‘setting/doing X’ has been presumed in the YAv. hapax (N. 28) *miiāzdi-*; however, this v.l. is now considered inferior and the form not *sprachwírklich*.

in IIr., in the present case we shall speak of “IIr. **máĭ(H)-as-* < IE **méĭ(H)-es-*” as a derivational basis, with no prejudices about the original meaning/root (and without positing an initial laryngeal, suggested neither by **mijās-* nor by **maĭa-* compounds [sūmaya- being read su° by Pp.!], the issue thus being largely irrelevant for the reconstruction of Iran. **mijās-*).

1.3.2. This allows to reconstruct following collocations with IIr. **máĭ(H)-as-* < IE **méĭ(H)-es-* (the issue whether from an old Seṭ basis **máĭH-as-* < **méĭH-es-* (← **méĭH*) or (be it only *argumenti causa*) from an old Aniṭ basis **máĭ-as-* < **méĭ-es-* (← **méĭ* ‘to strengthen’ remaining difficult, as per 1.3.1., s. also below):

- A. A (factive) verbal phrase IIr. **máĭ(H)-as-* + **d^haH* ‘to set/make a meal’ < (transponate) **méĭ(H)-os-* + **d^heh₁* ‘to set/make a strengthening/refreshment’, in Ved. *máyas* + *DHĀ* (middle diathesis in RV. 3,1,3; 8,39,4) ‘to set a refreshment/ meal for oneself’, ‘to make oneself / get a refreshment’; active: *máyas* + *KAR/kr̥*, RV. 1,31,7 *yás* [...] *máyah̄ kṛṇósi* ‘thou [...] who makest refreshment’, with the productive Ved. verb (HINTZE 2000: 67) for ‘to make’ (cf. RV. 8,20,24 *máyo na bhūta*, *mayobhuvan̄* ‘be refreshment for us, o you whose essence is refreshment’); this VP is nominalizable in (B.):
- (B.) A nominal phrase/juxtaposition based on (A.) is **mi(H)-ás(-s)* + root-noun **d^haH-* (*n. actionis*) ‘refreshment-setting, meal-offering’ > **mijás* + *d^hā-* > **mijáz* + *d^hā-*. Its first element can be posited, in accord with the reconstruction of the prevailing original paradigm of neutral s-stems as proterokinetic – strong stem nom./acc. R(é)-S(z)-D(z), weak stem obl. R(z)-S(é)-D(z) (SCHINDLER 1975) – as the original oblique stem **mi(H)-és-*, or directly as the orig. gen. sg. nt. **mi(H)-és-s*. The nominal phrase can continue (B_{set}) **mi(H)-ás(-s)* + root-noun **d^haH-* (with the old proterokinetic gen. sg. **mi(H)-ás-s* < **mi(H)-és-s* in its first element) [or (B_{aniṭ}) **m(i)ĭ-ás(-s)* + root-noun *d^haH-* ‘setting/making of meal’ (with a LINDEMAN variant of the old proterokinetic gen. sg. **m(i)ĭ-ás-s* < **m(i)ĭ-és-s* in its first element)], functioning as derivational/compositional base of (C.):
- (C.) A desyntagmatic formation **mi(H)ás-d^hH-a-*, adj., based on (A.)/(B.), is a compound / external derivative with suffix **-o-* (> IIr. **a*) > **mijás-d^ha-* > **mijáz-d^ha-*, meaning ‘related to the setting/making of (ritual) meal’. This adj. can continue (C_{set}) **miHás-d^hH-a-* [or (C_{aniṭ}) **mijás-d^hH-a-*, cf. (B.)], functioning as the base of (D.). If substantivized, such an adj. (C.) develops to:

- (D) **mijás/z-d^hHa-*, m. ‘the activity/situation related to setting/making of (ritual) meal’, i.e. simply ‘(ritual) meal-making’, in Ved. *miyédha-*, Av. *miiazda-*, m.:
- (E.) In a kind of *figura etymologica*, **mijás/z-d^hHa-* occurs together with Iran. **dā* < Iir. **d^haH* in collocations like **mijás/z-d^hH-a-m* + *d^haH* meaning ‘to sacrifice solid (ritual) meal’ < ‘to make (ritual) meal-making’ (see SADOVSKI 2018b: 179f. on *rādhas-DHĀ*, cf. *rādhas-KAR* in RV. 1,10,7, 8,64,1 *kṛṇuṣvā rādhaḥ* ‘leiste deine Sachleistung’; NARTEN 1986: 270f., and SADOVSKI forthc. on Y. 40,1 *mazdā ... mazdqm kərəšuuā* ‘o Mind-Setter, do/make (perform) [your (characteristic)] Mind-Setting!’), a periphrastic collocation with Iir. **kar* ‘to make’ + *n. act.* [etymologizing the name of *maz-dā* as *n. agentis!*]). This speaks for reconstructing Iran. *Dā* in such collocations as Iir. **d^haH* ‘to set/make’ (and not **daH* ‘give’). — S. also SADOVSKI forthc., for a similar derivational chain [A'] VP → [B'] NP → [C'] desyntagmatic formation in **-d^hH-a-* < **-d^hh₁-o-*, adj. → [D'] noun, in Ved. *médha-*, m. ‘meal offering’, perhaps also in [E'] *figurae (par)etymologicae* with *MAY/mi* [→ Ved. *mitá-medha-* ‘whose offering is supported/established’].

1.4. There are two further compounds attested in Iranian that probably contain **miiāzda-* as a second term, both being personal names:

1.4.1. YAv. (*)*Nərə-miiāzda-* (s. § 3) is inferred by the (patronymic) personal name *Nərə-miiāzdana-* (belonging to [> son of] *Nərə-miiāzda-*; the meaning of the derivational base of this patronymicon is to posit as ‘the one who has/makes/accomplishes the (solid) ritual meal [offering] of / proper to men/heroes’.

1.4.2. On the reconstructed OIran. **Dāma-miiāzda-* see § 4 below.

1.4.3. Within this semantic field, we should also refer to **Hṛta-d^hāma-* (§ 5).

1.5. As we have seen above, standard handbooks on Iranian onomastics deliver no further interpretation of the meaning or of the cultural context of this reconstruct; the only exception being the succinct reference in MAYRHOFER 1973: “vgl. Yasna 34,3 *myazdam ... ašāicā* [sic] ~ RV. 10, 70, 2 *ṛtāsya ... miyédah*”. Yet, a closer look into the texts shows that this is only the peak of the iceberg of a series of correspondences of formulae with the participation of these lexemes:

2. Of particular value are external phraseological contexts attested in the ritual poetry of the Younger Avesta and the Gāθās. The overt syntactical relations in such constructions are relevant for clarifying the internal syntax of the compound.

2.1. Thus, we find a series of collocations in which *aša-* and *miiazda-* occur together in constructions in the Young Avestan liturgy, formulaically repeated in texts of high ritual significance:

2.1.0. The *nivid-*/catalogues at the start of the (Long) liturgy (Vr. 1,3f–h) contain the genitival syntagma *ašahe miiazda-* ‘(solid) meal (offering) of Rightness’:

niuuāēḍaiiemi haṅkārāiemi
staotanāṃ yesniiānāṃ haṅdātanāṃ hufrāiiaštaṅaṃ,
ašaonāṃ ašaonināṃca ašahe miiazdanāṃ

I announce (the sacrifice / the *nivid-*catalogue of the invited deities)³, I organize/ arrange/convene ([the accomplishment of] the ritual) of the Staota yesniia put together and well-offered (in a solemn manner [KELLENS 2006-11: 1,26]) as of (**solid**) **ritual meals** of Rightness of/for the (male [deities]) full of Rightness, of/for the (female [deities]) full of Rightness.

This YAv. liturgical passage confirms the existence of the genitival phrase *ašahe miiazda-* ‘(solid) ritual meal of Rightness’. The skilfully arranged sequence of genitives with different syntactic function *ašaonāṃ ašaonināṃca ašahe miiazda-* also suggests that another genitival phrase, *ašaonāṃ miiazda-* ‘the (solid) meal (offering) of/for the righteous ones’, at least theoretically can be considered as an additional candidate for the interpretation of the compound. In both cases, we can start from a bahuvrīhi compound:

- If the underlying syntagma is *ašahe miiazda-*, such a bahuvrīhi would mean ‘whose *miiazda-* is the one of/to *aša-/Rta-*’: the relationship between the terms is either stative, ‘having/possessing the (solid) meal (offering) of/to R.’, or (rather) factitive, ‘making/accomplishing the (solid) meal (offering) of/to R.’
- In the case of an underlying syntagma *ašaonāṃ miiazda-* we could operate with a bahuvrīhi compound of the “double possessive” type described by SCHINDLER 1986: 394ff.: Here, instead of the possessive derivative *ašauuan-* ‘characterized with rightness’, the derivational basis *aša-* stays in its first term, in accord with the rule discovered by Wilhelm SCHULZE (1892: 39). The meaning of such a SCHULZE/SCHINDLER “double

³ Lit. “I invite down” resp. “I cata-logue”; cf. KELLENS 2017, on the Ilr. backgrounds SWENNEN 2015, with refs., and my paper at the *Max Planck* symposium in Berlin 2011 (SADOVSKI, forthc.).

possessive” would be, again, either stative, ‘having the (solid) meal (offering) of/for the righteous ones’, or (rather) factitive, ‘making/accomplishing the (solid) meal (offering) of/for the righteous ones’.

2.1.1. Important textual evidence that can influence the decision between these possibilities is contained in Y. 7,1a, in which the *offering of (solid) ritual meal* is mentioned at the very opening of the first stanza of the final Hāiti of the introductory lists located at the beginning of the Yasna liturgy:

※ *ašaiia dadqmi x^varəθəm mii-azdəm* (In accord) with Rightness, I place/set (i.e. I sacrifice) the (solid) meal [offering] as food.

In this last series of catalogic lists at the start of the Avestan liturgy, introduced by the “*ašaiia dadqmi*” sequence, the expression ‘I offer *miiazda-* [as solid food] (in accord) with / by (means of) *aša-*’ (*ašaiia dadqmi x^varəθəm miiazdəm*) is a constant formula in Y. 7,1 (cf. also Y. 7,20, 21, and – in the same prominent position at the start of the series – of Y. 8,1; on its continuation in Y. 8,2a s. below).

2.1.2. Another large textual series⁴ includes the sequence [*x^varəθəm miiazdəm*] within the formula *āiiese yešti* + dat./gen.obj.ind. ‘in a sacrificial manner / by (means of the) ritual I draw near / bring hither the (solid) meal [offering] as (solid) food, for satisfaction [*xšnūmaine*] of [+ list of gods]’. This group of formulae is characteristic of the litanies in Hāiti 3 of the Yasna, occurring in the Y. 3,1, 3,20; syntactically it is close to the scheme of Y. 7,1, too (cf. KELLENS 2015: 125): In both cases *x^varəθəm* and *miiazdəm* function as *acc. praedicati* and *acc. obiecti*, respectively, the corresponding predicates being *dadqmi* ‘I place/set’ in Y. 7,1 and *āiiese* (*yešti*) in Y. 3.

2.1.3. The third group of instances contains the formulaic expression ‘we both dedicate and assign these Haoma[-offerings]s and (solid) meal[-offerings]s and libations to + Dat.’. The construction sounds: [...] *imq haomq̄sca miiazdq̄sca zaoθr̄sca* (all in the pl.) + *pairi-ca dadmahī ā-ca vaēdaiimahī*, attested in Y. 4,1 and 4,3. It is resumed in Y. 24,1, after a topicalization of the verbal phrase *āuuaēdaiimahī* + Dat. + Acc. in formulae of the structure: [a] *ahurāi mazdāi haomq̄ āuuaēdaiimahī* ‘[it is] to A. M. [that] we assign these Haoma[-offerings]’, + posticipated acc. object including the commata [b] and [c] in the acc.: *imq̄ haomq̄sca, miiazdq̄sca zaoθr̄sca* ‘these Haoma[-offerings] and (solid) meal[-offerings]s and libations’ (Y. 24,1bc is also repeated in Y. 24,6).

⁴ See KELLENS 2015: 125f. (: Y. 3,1) and 2017: 67ff. on the varying structures in Y. 3, 4, 6, 7, with refs.

2.1.4. In Vr. 4,2, by this very formula a libator is assigned/consecrated (ā vaēdaiiamahi) to *miiazda-* itself, as a Ratu. In Yt. 8,1, the Ratu *miiazda-* is an object of solemn worshipping, introduced with the verb form [*frā.*]yazamaide.

2.1.5. Further combinations of *miiazda-* + verbs, esp. *DĀ*, follow up in § 2.4.3.

2.2. Of great value is the context of Y. 34,3, which offers the only *Gāṇic* attestation of *miiazda-*:

aṭ tōi ***miiazdəm*** ahurā
nəmanhā ašāicā dāmā
 gaēθā vispā ā xšaθrōi
 yā vohū θraoštā manahā
 [...]

Now [it is] to/for you, oh Ahura, **with reverence**, (and) to/for **Rightness**, [that] we **place/set** all the material/living beings, as the **(solid) meal [offering]**, in your power; (the beings) which you have ‘fattened’/nourished with Good Thinking [...]⁵

Here, we have the *dat. obi. indir.* of *aša-* within a factitive construction *gaēθā vispā miiazdəm dāmā* ‘we set all material (scil. *dāmā/-anī* ‘creation(s)’, *acc. obi.*) as *miiazda-* (*acc. praed.*) for R.’. Such a syntactic collocation implies a possible interpretation of the compound **(H)ṛta-miiazda-* as ‘who makes *miiazda-* to R.’. Phrasemes like *ašaiia dadami xvarəθəm miiazdəm* (2.1.) suggest an analysis of **(H)ṛta-miiazda-* as ‘who makes *miiazda-* in accord with / by means of *aša-*’

2.3. Vedic parallels of such collocations exist already in the RV. (cf. § 3.2.2.1):

<p>RV. 10,70,2: ā devānām agrayāvehá yātu, <i>nārāśamso</i> viśvárūpebhir áśvaiḥ <i>ṛtāsya pathā nāmasā miyédho</i>, devébh,yo devátamaḥ suṣūdat </p>	<p>As the top/lead driver of the gods, let Narāśamsa drive here with horses of every form(s);⁶ along the path of Rightness, with reverence, he will sweeten the (solid) meal [offering] for the gods, (being himself) the most divine of the gods.</p>
--	---

2.3.1. Such Old Indic instances demonstrate that not only the joint appearance of Av. *miiazda-* and *aša-* but the entire collocation consisting of the ele-

⁵ Cf. also HUMBACH – FAISS 2010: 100 and KELLENS 2013: 53–84, here 60.

⁶ Phrase division in accord with the hemistichion/pāda structure; diff. JAMISON / BRERETON 2014, but cf. Geldner 1951, ad locum: “Als der Götter Anführer soll Narāśamsa [...] kommen. Auf dem rechten Wege unter Verbeugung soll der Göttlichste das Opfermahl für die Götter zubereiten.”

ments [SETTING] + [(SOLID) RITUAL MEAL] + [WITH REVERENCE] + [TO / (ON THE PATH OF) / WITH RIGHTNESS] has formulaic character and a good chance to have been inherited from IIr. ritual rhetorics (cf. 2.2., 3.2., 4.2., 5[1.3.]).

2.3.2. All three components that we saw in the Gāthic quotation Y. 34,3 in § 2.2. – *Aša-*, *nəmah-*, and *myazda-* – fully correspond to the Vedic terms from § 2.3. – *Rtá-*, *námas-*, and *miyédha-*. The added value is their *joint occurrence!*

2.4. By this testimony, the initially hypothetical interpretation of IIr. **H₁ta-mi(H)azda-* as a factitive bahuvrīhi compound with *miiazda-* ‘(solid) meal [offering]’ as second term becomes more probable, being now based on lexical combinatorics proven not only within Avestan but also in the Vedic context:

2.4.1. The meaning of the compound suggested by the combinatorics of its terms as lexemes in free syntactic contexts ranges from ‘having/accomplishing *miiazda-* of/to Rightness’ – cf. § 2.1.0. *ašaonəm ašaoninəmca ašahe miiazdanəm* ‘of (solid) meal[-offering]s of/to Rightness of/for the righteous ones (m./f.)’ – via ‘accomplishing *miiazda-* to Rightness’ (§ 2.2.) up to ‘having/accomplishing *mii-azda-* (in accord) with / by (means of) Rightness’, as per § 2.1, 2.3.

Typically, a bahuvrīhi compound nominalizes these variant ranges into a unitary formation that allows for a plurality of interpretations, faithfully preserving, however, the religious constants contained in the corresponding formulae.

2.4.2. The factitive interpretation is suggested by the fact that in the prominent contexts of its attestation, *miiazda-* appears as effected object of the verb *DĀ* ‘to set/accomplish/make’. – In spite of the limited character of its attestations (RV. only 7×), Ved. *miyédha-*, too, the etymological and ritual counterpart of *miiazda-*, shows similar usage: with *āsāna-*, the regularly formed med. part. of *ĀS* (Werba 1997: 449f.; < transponate **h₁eh₁s-ṛh₁no-*) attested only twice in the RV., both in Maṇḍala VI (here and 6,106b). (The expression ‘with the [solid] meal[-offering]s set [here]’, indirectly, also speaks in favour of the interpretation of the Iranian root *DĀ* in *miiazdam* + *DĀ* not as ‘to give’ but as ‘to set’):

RV. 6,51,12:
 nú sadmānaṃ div yāṃ nāmśi
 devāḥ,
 bhāradvājaḥ sumatīm yāti hótā |
 āsānébhir yájamāno miyédhair,
 devānāṃ jánma vasūyúr vavanda
 ||

JAMISON / BRERETON 2011: 2,847:
 To attain the one whose seat is in heaven,
 o gods,
 the Hotar of the Bhāradvājas now begs for favor.
 Sacrificing with the (solid) meals set here,
 desiring goods, he has extolled the breeds
 of the gods.

2.4.3. The Avestan formulae of 2.4.2. appear, for instance, in a series of repeated ritual invocations contained in the Young Avestan *Āfrīnagān*-i *Gāhānbār*. The cycles of hymns belonging to the liturgical genre of *Āfrīnagān* are known to contain archaic ritual phraseology:⁷ starting already with the concept of Av. *ā-frī-* that goes back to the cultic technical sense of Ilr. **PRĪ* (± *ā*) ‘to satisfy by cultic activity’, and the corresponding Ved. genre of liturgical hymns *ā-prī-*. Notably, the archaic term Av. *mīazda-*, later substituted, e.g., by the concept of *draona-/drōn* in Iranian,⁸ (just as Ved. *miyédha-* ‘[solid] ritual meal [offering]’, attested only 7× in RV., is increasingly recessive after the RV. period), is most frequently used in the Young Avestan *Āfrīnagān*, cf. the following selection:

gen. sg. <i>mīazdahe</i>	8× of 11× in A.	A. 3,7.8.9.10.11.12; A. 8,12; N. 72.81.83
Derivatives:		
nom. sg. <i>mīazdauuā</i>	6×, only in A.	A. 3,8.9.10.11.12; A. 8,12
gen. pl. <i>mīazdauuanqm</i>	3	N. 61–63
neg. <i>a-mīazda-uan-</i>	7×, only in A.	A. 3,7.8.9.10.11.12; A. 8,12

2.4.3.1. With the formula *mīazdəm DĀ*, A. 3,3 asks: *dātāca aēte mazdaiiasna ... mīazdəm* ‘And set/make, o Mazdaiiasnians (present) here ... (the) *mīazda-*’, Vd. 8,22 prescribes: *mīazdəm daiθiārəš* ‘they might set/make the *mīazda-*’, FrK 72[f.] describes: *mīazdəm daḍāiti* ‘one sets/makes the *mīazda-*’ (both times: ‘*gaomaṇtəm maḍumaṇtəm* ‘which contains “cow products” and wine’);

⁷ On the interconnections between the Avestan *ratu-* catalogues, esp. *Visprad* lists from the Long Liturgy but also from *ā-frī-* rituals, on the one hand, and the Vedic *ṛtviya-* catalogues, esp. in lists from the *ṛtu-yājā-* liturgy but also from *ā-prī-* rituals, cf. SADOVSKI 2018a, with numerous details: the parallel structures in the Veda range from RV(Kh) traditions of archaic liturgy up to popular rites. Thus, invocations of the *ṛtu-*s in the ritual of the *pitar-*s (KRICK 1982: 40 with n. 88 and lit.) show parallels with the Avestan idea of *ratu-frī-* ‘the satisfaction of the *ratu-*s’ esp. in the cult of the Avestan *frauuāṣi-*s. See also SADOVSKI 2018c: 312–314 on the Vedic *vaiśvadeva-* ritual (BaudhGS 2,8) with its 25 oblations offered in the middle of the house (cf. GONDA 1983: 29) with invocation formulae addressed to the 33 [!] *Vāstoṣpati*, protectors of this demiurgic act, as a parallel to the Avestan list of [33!] divine entities, protectors of creation, within the litanies of Yasna 71.

⁸ Here we should not forget, too, that the ritual Y 60–61, being the Second *Drōn* after Y. 59, is explicitly dedicated to *Dahma- Ā-frī-ti-*.

2.4.3.2. and A. 3,6 affirms: “xšaθrəm [...] aibī dadəmahī [...] mazdāi ahurāi ašāicā vahištāi: dātō hē **miiazdō ratufrīš**” ‘we “set”/assign Power to Ahura Mazda and Aša Vahišta. Accomplished by him is the **miiazda-** that satisfies the Ratus’. – The technical term ratu-frī(ti)- contains the same second term of cultic satisfaction of (divine/ritual) authorities as in Av. ā-frī- Ved. ā-prī-; see n. 7, 8.

2.4.3.3. One of the most frequently occurring versions of the *miiazdām* DĀ formula occurs in Āfrīnagān-i Gāhānbār, too: in the long sequence in A. 3,7–12 describing rituals and prescribing punishments for not fulfilling them properly:

A. 3,7:
 pañcāca caθβarəsātəmca maiḍiiōi.
 zarəmaiehe
ašāhe vahištahe daθušō:
disiāt hē aṅhe auuāt mīzdām parō.
 asnāi aṅuhe
 yaḍa aētahmi aṅhuuō yaṭ astuuaiṅti
 hazarəm *maēšinam daēnunam
 paiti.puθranam
narām ašaonam ašāiia vaṅhuia urune
para.daiθiāt
 aēuuahē hātəm cinmānahe
 yaṭ ašāhe vahištahe.:
 paoiriieheca nā **miiazdahe** aḍāiti
 spitama zaraḍuštra
 yaṭ maiḍiiōi.zarəmaiehe
miiazdauū ratuš **amiiazdauuanəm**
 ratunaēm
 aiasnīm daste aṅtarə mazdaiiasnāiš

[There are] forty-five [days up to the (main) day] of Maiḍiiōi.zarəmaia, [in the month of the Best Rightness, [on the day] of Him-who-has-set-up (=Creator): Such a reward for it will be assigned to him (= sacrificer) for the future live, as (if) he would have given over / donated in this life, the material one, thousand sheep-mothers together with the[ir] young (calves), for the uruuan-soul – in accord with good Rightness – of righteous men, of (some)one of the being ones wh[o] “sets/puts on” / cares of the Best Rightness. And for not-accomplishing the first **miiazda**, o Sp.Z., the one of Maiḍiiōi.zarəmaia, the Ratu who [accomplishes] the **miiazda** declares the [layman] under (this) Ratu’s guidance who has no[t accomplished the] **miiazda**, unable (to participate) in the Yasna among the Mazdaiiasnians (cf. Wolff 1910: 310, HINTZE 2000: 316f. vs. MILLS 1886: 369).

The formula in A. 3,7 is repeated with variations of the ritual situation in st. 8–12, but always in this constant shape. It contains a series of interesting enlargements of the conceptual sphere characterizing the *miiazda-* offering ‘in accord with Rightness’: For accomplishing a *miiazda-*, a reward is foreseen, designated as *mīzda-*; this shows a conscious (par)etymological word-play of *mīzda-*, Ved. *mīdhá-/mīlḥá-*, with *miiazda-*, Ved. *miyéḍha-*, both sharing at least

their second term and formation typology (s. § 1.3.; on the first term SA-DOVSKI forthc.).

2.4.4.1. In both the Avestan and the Vedic liturgy, we find the most important technical terms of IIR. in clusters which often offer amazing phrase-by-phrase correspondences. – For the Avesta, cf. Vr. 11,2:

<u>ima haoma</u> ima haomiia	<u>Here (are) the Haoma</u> [-offering]s and the (instruments for [preparing]) Haoma
<u>ima starəta ima miiazda</u>	<u>here the spread</u> [mats], here the [two] (solid) meal [-offering]s,
ima asma paoiriia <u>dāmañ</u>	here the stone[-sky], the first (in the ²) <u>creation</u> [...]
[...]	<u>here the ritual grass</u> (barəsmān-) <u>strewn (in accord) with Rightness.</u>
<u>imat barəsmā ašaiia fra-</u>	
<u>starətam</u>	

Its excellent Vedic counterpart (not really studied so far) is RV. 1,177,4:

ayám yajñó devayá ayám miyédhañ- ,	Here is the sacrifice , proceeding to the gods; here the (solid) meal [offering] , <u>here the soma</u> , Indra, here the sacred formulations.
imá bráhmāñy <u>ayám</u> indra <u>sómah</u>	<u>The ritual grass</u> [is] <u>strewn</u> ; drive forth toward it here, able one;
<u>stīrnám barhír</u> á tú śakra prá yāhi,	drink, having sat down on it, unharness your two fallow bays here (after JAMISON / BRERETON 1,377).
pībā niśádyā ví mucā hári ihá	

Cf. RV. 3,32,12, where we observe the same pairwise appearance of *yajñá-* and *miyédha-*:

yajñó hí ta ind _a ra várdhano	Because the sacrifice has become your strengthener, Indra, and the dear (solid) meal [offering] of
bhūd,	pressed soma also [...]
utá priyáh sutásomo miyédhañ	

2.4.4.2. It is the libator priest, Ved. *hótar-*, Av. *zaotar-*, to organize the *miiazda-*. Thus, N. 54 (72) attests: †ciš zaotarš kairim aṇhaṭ **miiazdahe* **aiiqn?* **g[āḍ]āsci frasrāuuaiiāiti* ‘What is the *zaotar*’s duty on the days of the [solid] offering? [...] He shall recite the Gāthās’. And in RV. 3,19, the cyclic composition (1a) agním hótāram prá vṛne miyédhe ... (5a) yát tvā hótāram anājan miyédhe ‘I choose Agni as the **Hotar at the (solid) meal [offering]**’ ...When the gods will anoint you as the Hotar at the (solid) meal [offering] emulates a ritual of priest election of IIR. age!

2.4.5. The sequence A. 3,7–12 also contains two derivatives of *miiazda-*:

2.4.5.1. the possessive/factitive **-uēn-* derivative *miiazda-uuan-* ‘the one who has/accomplishes *miiazda-*’ and a further derivative of this **-uēn-* possessive:

2.4.5.2. the negative *amiiazdauuan-*. The 6 attestations of the latter exclusively appear in A. 3,7–12; the same is valid of the 6 occurrences of the nom. sg. of *miiazda-uuan{t}-*, while we find the acc. sg. *miiazdauuanəm* in N. 61–63.

2.4.5.3. Yt. 13,64 contains another YAv. derivative, *miiēzdin-* ‘one who has/accomplishes *miiazda-*’: here, the *ašāunəm frauuašaiiō* ‘*Frauuašis* of the *righteous* ones’ ride by ten thousands among those offering (solid) meals (*miiēzdinqm*).

2.4.5.4. To finish the survey of derivatives, Ved. *miyedhya-* ‘related to / participating in the (solid) meal[-offerings]’ is attested only in voc. and only with the voc. *agne*, in RV. 1,36,9, RV. 1,44,5, and RV. 1,26,1.

2.4.5.5. Notably, Y. 8,2 states that men have ‘won’ the *miiazda-* *with Rightness*:

<i>x̣^oarata narō aētəm miiāzdəm</i>	<i>Swallow/eat, men, this (solid) meal[-offering]s</i>
<i>yōi dim hañhāna ašāca</i>	which you have won [both] <i>with Rightness</i> and
<i>frāretica</i> ✱	with ‘propagation of Dawn’ ⁹

This passage not only confirms the factitive reading of **H₁ta-mi(H)azd^hHa-* but explicitly presents men/heroes (*nar-as*) as those who fulfil/consume *mii-azda-*. At this point, having examined the contexts of Av. *miiazda-* with its derivatives, let us turn to the other compounds related to this phraseological domain: the two names with **mi(H)azda-* as second term and one with **(H)₁ta-* as first term:

3. Compounds with *miiazda-* as *Hinterglied* in Avestan might be rare but they do exist: Thus, a formation *Nərə-miiazdana-* is attested in Yt. 13,110, in the commemorative list of the first members of the Zoroastrian community:

3.1. YAv. *Nərə-miiazdana-* (IPNB 1,65 #233; SCHMITT 2003: 367ff.) is analyzable as a patronymic formation in *-ana-* ← a PN **Nərə-miiazda-* with the meaning ‘the one who has/makes the (solid) meal [offering] “of” / proper to men’.

⁹ On possible allusions to the context of ritual pragmatics, cf. KELLENS 2006–2011: 2,12. For the ritual context of the *Srōš Drōn* (as compared with the *Drōn Yašt* / *Bāj Dharnā*) s. Redard 2015: 199f.

3.2. The analysis of the compound gives us the occasion to explore the phraseological combinatorics of the lexemes that form its terms:

3.2.1. Av. *miiazda-* [+ *ašca-*] + *nar-*: a phraseme apt for comparison is again Y. 8,2:

<i>x^oarata narō aētəm miiazdəm</i>	Swallow/eat, <i>men</i> , this (solid) <i>meal</i> [<i>offering</i>]
<i>yōi dim harjhāna ašca frərətica</i> ※	that you have won [both] <i>with Rightness</i> and with 'propagation of Dawn'

3.2.2. Ved. *miyédha-* [+ *rtá-*] + *nar-*: MAYRHOFER 1979: 65 #233, builds a bridge between 3.2.1 and the word-family of Ved. *médha-* 'offering (of solid and liquid meal)' and the PN Ved. *Ṛ-médha-*, admitting, however: "[V]ed. N[Ṛ-*médha-*] wird von [Rüdiger] Schm[itt] bei Schle[rath 1966: 2,]164 allerdings in einen anderen Zusammenhang gestellt (: V[d.] 18,51 *narəm ... maqzdrəm*)".

3.2.2.1. In the RV. stanza already quoted above, the *miyédhas-* [n.!] is offered by *Nārāśamsa-*, the deified performer of the 'praise-songs of men', incorporating the heroic poetry and protecting male elites: both of the Iir. establishment (kings, warriors) and of the ephebic age-groups (*Männerbund*);¹⁰ RV. 10,70,2:

<i>ā devānām agrayāv_{ai} há yātu,</i>	As the lead driver of the gods, let <i>Nārāśamsa</i>
<i>nārāśamsō viśvārūpebhīr āsvaiḥ </i>	drive here with horses of every form(s);
<i>rtásya pathā nāmasā miyédho,</i>	<i>along the path of Rightness, with reverence,</i>
<i>devébhyo devātamaḥ suśūdat </i>	for the gods, the most divine of the gods will sweeten the (solid) <i>meal</i> [<i>offering</i>].

3.2.2.2. Cf. RV. 7,36,5, where the notion of 'fortifying nourishments' is close to the idea of *miiazda-/miyédha-*; the rites are said to be performed by men/heroes, in a hymn dedicated to Rudra, the main deity of the OInd. *vṛātyas* (*Männerbund*):

<i>yājante asya sakhyām váyaś ca,</i>	The <i>reverent</i> ones worship his fellowship and
<i>namasvīnaḥ s_u vá rtásya dhāman </i>	vitality on the <i>domain of Rightness</i> itself;
<i>vī pṛkṣo bābadhe ṛbhi</i>	he has thrust outward the fortifying nourishments
<i>stāvāna_h-,</i>	when being praised by <i>men</i> . This <i>reverence</i> is
<i>idám nāmo rudrāya pr_{āyī}stham </i>	dearest to Rudra (cf. JAMISON / BRERETON).

¹⁰ E.g. OLDENBERG 1967: 210–224; OBERLIES 2012: 74, 155; SADOVSKI 2018a: 92f., 125; 2018c: 321f.

WEISS 2009 links Iir. “(mi̇jaz-)d^ha-” to Umbrian **eru-fo-s* ‘part of sacrifice partaken of by *men*!’ — What is of high relevance, is the localization of this veneration ‘in the setting of Rightness itself’, also with regard to the next two compounds:

4. In the corpus of (possibly) Iranian names in New/Late Babylonian transmission, the excellent dictionary of Ran ZADOK (2009: 153, #233) lists the male name <[D]a-²-ma-^rmi-as/z-ta>, <[D]a-²-ma-mi_j-as/z-ta>, m., with a detailed dossier of its attestation and prosopography. The name origins from Nippur, the Muraššû- archive, the source is dated 16.VII.41. Art. I. = 30.10.424 BC.

4.1. As IPNB 7/1B, 153 (with refs.) underlines, earlier interpretations do not convince for phonological or for morphological and semantical reasons: this concerns Eilers’ **Dāmi-Mazdā* ‘Schöpfer Mazda’ and Hinz’s **Dāma-myasta* ‘„Paradies-vereint“ (reading: -mi-as-ta), respectively. ZADOK himself hesitates in assessing the name: “Unsicher. ... Eine Verbindung des Namens mit **m(i)jazda* (av. *mīazda*) „Opfergabe“ ist erwägenswert; aber da der Name sich vielleicht auf einen Mannäer bezieht, könnte er auch nicht-iranisch sein.”

4.1.3. If the formation is Iranian, its reconstruction will be as a compound OIran. **dāma-mijazda* < Iir. **d^haHma-mi(H)azd^ha-*, with a first term Iir. **d^haHman-* ‘seat, location, setting (esp. of a supernatural phenomenon)’ > Iran. **dāman-*, GA. ‘id.’ but YAv. also ‘creation [*n. actionis* or *n. concretum*,] creature’, and a second term Iran. **mijazda* < Iir. **mi(H)azd^ha-* ‘(solid) ritual meal [offering]’.

4.2. Possible meanings of this [reconstructed] Iranian compound are:

4.2.1. as a bahuvrīhi ‘whose (solid) meal offerings are (for/to/of/in/by) **d^haHman*’, for word-formation and onomastic reasons much more probable than

4.2.2. as a tatpuruṣa: ‘meal (offering) of (for/to/of/in/by) **d^haHman*’.

4.2.3. Another theoretical interpretation would be as a ritualistic sentence-name based on a Gāḍic phrase. Parallels of this not unusual Avestan name type can be:

(A.) the YAv. Saošiiant- names Yt. 13,129: *Astuuat.ərətā-* (Malandra 1971: 145 + comm. 215, IPNB 1,22f. #36, MAYRHOFER 1977: 13 + n. 27), derived ← Gāḍic liturgical text *astuuat ašəm xīiat* ‘may Rightness be osseous/material[ized]’; on names (B–D) as derivatives from mere Av. text strings see MAYRHOFER 1977: 12f., SCHMITT 2003: 365, all three in Yt. 13,120:

(B.) *Ašəm.yeŕjhe.raocā* m. (IPNB 1,25 #49) ← linear string Y. 12,1 *yeŕjhe ašəm yeŕjhe raocā*;

(C.) *Ašəm.yeŕjhe.varəza* m. (IPNB 1,25 #50) ← a string not transmitted so far.

(D.) *Ašəm.yahmāi.ušta* m. (IPNB 1,25 #49) ← [*ašəm* +] *yahmāi.ušta* in Y. 43,1.

If **Dāma-myazda-* would belong to the same type, it could be analyzed as a *reprise* of the formula “We set/deposit a (**solid**) **meal** [*offering*]” and traced back to the *Ahunauvaitī Gāθā*, as a quotation name with a clear interpretation in a Zoroastrian context, in particular within the (meta-)ritualist system of the *Gāθās*.

4.3. The first term, Ilr. **dháH-man-* is a fundamental religious notion in Ilr.:

4.3.1. Usages of Ved. *dháman-* suggest interpretations like *nomen actionis/abstractum/concretum* ‘setting’ (*Setzung*, OLDENBERG 1967: 352–362, esp. 353ff.), i.e. both factitive ‘setting, establishing, arranging/–ment’, as ‘creating/–tion, instituting/–tion’, and stative ‘setting, *Einstellung*; set-up, *Satzung*, constitution; (divine) disposition’, as concr. ‘establishment, institution, creation, *domain*’ (GONDA 1967:) ‘location of divine presence, emanation of a supernatural force, numen’, even (invisible but perceivable) ‘form/manifestation of divine setting/disposition’.

4.3.2. Iran. **dāman-*: GAv. *dāman-* shows a similarly complex semantic range, while YAv. *dāman-* prevalingly means ‘creation/creature(s)’ (*AiW* 734–736).

4.4. Again, phraseology and contextual knowledge of sacred texts are of decisive value for the choice between these interpretative options:

4.4.1. YAv. *dāman-* mostly combines with *vispa-* or is substituted by it (s. 2.2):

– Y. 71,10, Vr. 12,4: *vispe ... varjuhiš dāman ašaoniš ... yāiš dadāθa* ‘all these good creatures ... full of Rightness, {by} which you have created [...]’; – Y. 71,6 *vispāča dāma mazdaθātā ašaoniš* ‘and all creatures, mazdā-given, full of R.’ – Y. 1,16, Y. 3,18, Y. 4,21 etc. *vispanamca spəntahe mainiiūš dāmanam, ašaonam ašaoninamca, ašahe raθβam* ‘and of all creatures of Spənta Mainiiuu, the (male) ones full of R. and the (female) ones full of R., Ratus of R.’; – Y. 16,1: *ahurəm mazdām ... dātārəm vohunam dāmanam* ‘A.M., the creator of good creatures’; – Yt. 8,48: *vispāiš ... yāiš spəntahe mainiiūš dāman* ‘by all the ... creatures of Sp.M.’; cf. Yt. 10,142, Yt. 12,3, Vyt. 51 (without *vispāiš*), V. 5,33, V. 13,28, etc. (*dāmanam*, without *vispāiš*); – Yt. 13,93: *vispā spəntō. dātā dāman* ‘all Sp.M.-created creatures’.

4.4.2. All ‘ingredients’ of our compound and its contexts are contained in Y. 34,3:

<p>Y. 34,3: at̥ tōi miiazdəm ahurā nəmanhā ašāicā dāmā gaēθā vispā ā xšaθrōi yā vohū θraoštā manahā [...]</p>	<p>Now [it is] to/for you, oh Ahura, <u>with reverence</u>, (and) to/for <u>Rightness</u>, [that] we <u>place/set/deposit</u> all the material/living (creation[s] [<u>dāmā/dāman</u>]), as the (solid) meal [offering], in your power; (the beings) which you have ‘fattened’/nourished with Good Thinking [...]¹¹</p>
---	---

4.5. Since GAv. dāmā, YAv. dāma formally coincides with the acc. sg. of dāman-, n. ‘creation, being’ (pl. *qan, GAv. *anī), in the same formula dāmā (verb, 1. pl.) might also have secondarily been brought into connection with the homonymous GAv. dāmā – as if the whole were a nominal sentence: ‘To you, oh Ahura, (and) to Rightness, with reverence, [we bring] the creation [*quasi acc. obi.*], all the material beings [*apposition*], as **(solid) meal [offering]** [*acc. praed.*] in your power, etc.’. dāmā + gaēθā vispā, being in immediate contact position beyond the hemistichion caesura, form a string stimulating the contextual creation of such associative chains.

4.5.1. We should not forget that if strings of word-forms only loosely or not at all connected with each other can reach the status of intellectually/religiously conditioned names [s. 4.2.3., *Ašəm.yeŋhe.raocā*], just because these strings make part of the sacred texts of the Avestan liturgy, the chance to allude to Zaratustra’s *verba ipsissima* from the Gāθās in a name could have been even more appreciated by later Zoroastrians.

4.5.2. Such an association additionally facilitates the spread of a primarily *sprachwirkliche* bahuvrīhi formation *Dāma-miiazda- ‘whose (solid) meal [offerings] are (the) beings’, ‘who has/brings meal[-offering]s (consisting of all) the beings’ (rather than [similarly possible!] ‘whose (solid) meal[-offering]s are for the creation’ [phraseology unattested] and instead of *Vīspa-miiazda-/Gāēθa-miiazda-).

5. At any rate, the compound *Dāma-miiazda- is clearly based on linguistic reality, as shown by the phrasemes in 4.3.2 and by the usage of Iir. *dhāH-man- as compositional term in other lexemes that nominalize phrases with lexemes of the ritual/religious sphere, one of which I would briefly cite here:

¹¹ Cf. also HUMBACH / FAISS 2010: 100 and KELLENS 2013: 53–84, here 60.

5.0. In the corpus of Old(est) Indic, probably already Ilr., names of Mitanni rulers, there is the well-known name *Ar-ta-ta-ma*, m., interpretable as ^{*(H)}*ṛta-dāman-* < Ilr. ^{*}*Hṛta-d^haHman-*. Its attestation and the respective prosopography have belonged to the objects of most active research in the second half of the 20th c.: see MAYRHOFER 1987 (with refs.), more recently e.g. RAULWING 2013.

5.1. Possible interpretations of the compound are:

5.1.1. as a substantival tatpuruṣa-: ‘setting/domain/creation(s) of Rightness’.

5.1.2. as a (substantivized) bahuvrīhi-: ‘(one) whose setting/domain/creation is (in/by/ for/[in accord] with...) Rightness’, ‘who “has” the setting/domain of R.’.

5.1.3. Another theoretically possible underlying structure works with a *n. actionis* ‘who accomplishes setting/creation (in accord) with / by means of R.’; for the corresponding VP, cf. the juncture ‘I set X as Y (in accord) with R.’ (2.1.1.).

5.1.4. Since the spelling of the Mitanni name does not allow for any conclusions regarding its accentuation, we have to look for appellative correspondences:

5.2. In Vedic, *ṛtá-d^hāman-* is attested twice, both times in the Vājasaneyi-Saṃhitā, namely at VS. 5,32 (identificational epicleresis) and 18,38 (cultic epithet).

5.2.1. In both attestations, the first-term accent of *ṛtá-d^hāman-*, nom. sing. *ṛtá-d^hāmā*, clearly suggests an interpretation of the compound as a bahuvrīhi:

VS. 5,32h-i (ed. WEBER 1849: 151):
mṛṣṭò^ṽ a^ṽ si havyasúdanah |
ṛtád^hām_{ṽa} si s_ṽ vārjyotiḥ ||

VS. 18,38 (ed. WEBER 1849: 577):
ṛtāsád **ṛtád^hām**_{ṽa} gnír
gand^harvás tásyáušad^hayo
'psarásó múdo náma |
sá na idám bráhma kṣatrám
pātu
tásmai s_ṽ vāhā vāt^ṽ táb^hyaḥ
s_ṽ vāhā ||

Griffith 1899: 41:

[...] Swept clean art thou, preparer of oblations.

Thou art the seat of Law, heaven's light and lustre.

Griffith 1899: 166:

Maintainer of Law, true by nature, Agni is the Gandharva. The plants are his Apsarases, namely Delights.

May he protect this our Priesthood and Nobility.
To him All-hail! Ave! To those All-hail!

5.2.2. Consequently, against the classical translation delivered by Griffith, Pada *i* of VS. 5,32 has to be interpreted as ‘Thou art the one whose setting/domain is (in/by/with etc.) Rightness, the one who has the light of sunshine’. The suggested translation for 18,38a is, correspondingly: “Being one who sits/dwells in Rightness, one whose setting/domain is (in/by/with etc.) Rightness, Agni [...]”.

5.3. The phraseology of **d^háH-man-* in IIr. sacred poetry shows relevant contextual combinatorics with semantic consequences for the analysis of the compound:

5.3.1. Iranian displays evidence for syntagmata containing both lexemes:

5.3.1.1. There is the theoretical possibility to interpret **Hrtá-d^haHman-* as a double possessive compound according to SCHULZE/SCHINDLER’s rule (above, 2.1.0.), of the meaning ‘whose setting/establishment/domain is righteous’, in YAv. esp. ‘whose creation(s) are righteous’: Such a compound would nominalize syntagmata of **d^haH-man-* + possessive adj.s like *dāman- ašauuan-* ‘righteous creation, creature(s) full of Rightness’, syntagmata well-attested in YAv. texts:

- Vr. 19,1f. ... *dāman ašauuanō yazamaide* ‘... we worship the creations, the ones full of R.’;
- Vr. 7,4abc *auuā dāman ašauuanō yazamaide yā hənti paoiriiō.dāta paoiriiō.fraθbaršta ašnāatca apāatca* ... ‘we worship those creations which are first/primordially created, primordially shaped, before both heaven and water’; cf. also Vr. 19,2, Yt. 6,2 (*buuat dāma ašauua yaoždāθram*); Y. 19,19.
- Y. 16,3, Vr. 11,2, Vyt. 43: *pauruuā dātā dāman ašaoniš’ daθušō* ahurahe mazdā [...] *yazamaide* ‘we worship the first/primordial creations of A.M., Him-Who-has-created(-all)’;

5.3.1.2. In GAv., Y. 48,7 states: *ašā ... at hōi dāman θbahmī ā dām ahurā* ‘[protected/covered] by/with Rightness ...; its/his settings/domains [are] in thy house, o Ahura’. If *hōi* refers not to the man protected or the thought covered by/with Rightness but to Aša it-/himself (topicalized at the beginning of 7c), this locus will speak in favour of an underlying **Hrtásja d^háH-man-*. The corresponding bahuvrīhi will mean ‘having/making creations (full) of Rta’.

5.3.2. Ved. *ṛtásya dhāman-* ‘setting/domain of Rightness’; concr.: ‘creation(s) of Rightness’, occurs in artfully shaped stanzas such as:

RV. 1,43,9

yās te prajā amṛtasya
parasmin dhāman rtasya |
mūrdhā nābhā soma vena
ābhūṣantiḥ soma vedaḥ ||

The creatures [=gods] that belong to you, the immortal one—as their head you seek them in the highest domain of Rightness [=heaven] and in the navel (of the earth?) [=ritual ground], Soma. You will know them as the ones who attend upon (you), Soma. (modif. after JAMISON / BRERETON 1,153)

RV. 7,36,5:

yājante asya sakhyāṁ váyaś ca,
namasvīnaḥ s vá ṛtásya dhāman |
vípṛkṣo bābadhe nṛbhi stāvāna,
idāṁ nāmo rudrāya pr stham ||

The reverent ones worship his fellowship and vitality on the domain of Rightness itself. He has thrust outward the fortifying nourishments when being praised by men. This reverence is dearest to Rudra (JAMISON / BRERETON 2,930, with modif.)

Cf. RV. 4,7,7:

sasāya yād víyutā sāsminn
ūdhann
ṛtásya dhāman raṇáyanta
devāḥ |
mahám agnir nāmasā
rātāhavyo
vér adhvarāya sādām íd
ṛtāvā ||

When, at the separation of the grain, at the same udder [=the sacrifice?], on the domain of Rightness, the gods take pleasure, great Agni, on whom the oblation is bestowed with reverence, the righteous one, is always pursuing (his missions) for the ceremony (cf. JAMISON / BRERETON 1,570, with modif.).

as well as the collocations in RV. 10,124,3b ṛtásya dhāma ví mime purūṇi ‘I measure out the many domains of Rightness’ and, discontinuous, RV. 1,123,9: ṛtásya yóṣā ná mināti dhām, ‘she does not violate the domain of R.’. These formulaic attestations of ṛtásya dhāman- favour an interpretation of the compound *(H) ṛtā-dāman- < Ilr. *Hṛtá-d^haHman- as ‘whose setting/domain is (the one) of R.’.

Phraseological comparison as in § 1–5 demonstrates how the study of the syntactic structures and poetic collocations in the hymns of the Avesta and Veda turns out to be seminal for our understanding of archaic compounds – epithets and names – which occur not only in sacred texts but also in frameworks of everyday life such as the voluminous catalogues of personal names in the multilingual lists attested in the Persepolis archives and all over the Great Kingdom of Achaemenid Persia.

Abbreviations

- | | |
|--|--|
| (I.) (P)IE = (Proto-)Indo-European. | (III.) Avestan: |
| - (P)Ir. = (Proto-)Indo-Iranian. | (a.) GAv. = Gāthic Avestan, esp.: |
| - (Hom) Gk. = (Homeric) Greek. | - Y. = Yasna. |
| (II.) Ved. = Vedic: | - YH. = Yasna Haptaṅhāiti. |
| (a.) RV. = Ṛgveda-Saṃhitā. | (b.) YAv = 'Young(er)' Avestan, esp.: |
| (b.) AV. = Atharvaveda-Saṃhitā, esp.: | - Y. = (Younger Avestan) Yasna. |
| - AVŚ = Atharvaveda-Saṃhitā, Śaunaka branch. | - Yt. = Yašt. |
| - AVP = Atharvaveda-Saṃhitā, | - Vd. = Vidēvdād. |
| Paippalāda branch. | - Vr. = Vīsprad. |
| - Kauś = Kauśika-Sūtra. | (IV.) Bibliographical abbreviations: |
| (c.) YV = Yajurveda, esp.: | - EWAia. = MAYRHOFER 1992–2001. |
| (α.) Black YV: | - IPNB = <i>Iranisches Personennamenbuch</i> . |
| - TS. = Taittirīya-Saṃhitā. | - IPNB 1 = MAYRHOFER 1979. |
| (cβ.) White YV: | - IPNB 7/1B = ZADOK 2009. |
| - VS. = Vājasaneyi-Saṃhitā. | - JBr. = JAMISON and BRERETON 2014. |
| - ŚB. = Śatapatha-Brahmaṇa. | |

References

- BENVENISTE 1966
 BENVENISTE, É.
 1966 *Titres et noms propres en iranien ancien*. Paris.
- GONDA 1967
 GONDA, Jan.
 1967 *The Meaning of the Sanskrit Term Dhāman-*. (Verhandelingen der Koninklijke Nederlandse Akademie van Wetenschappen. Afdeling Letterkunde. N.R., 73/2). Amsterdam.
- HALLOCK 1969
 HALLOCK, Richard T.
 1969 *Persepolis Fortification Tablets*. Chicago, IL.
- HINTZE 2000
 HINTZE, Almut.
 2000 *„Lohn“ im Indoiranischen. Eine semantische Studie des Rigveda und Avesta*. (Beiträge zur Iranistik 20). Wiesbaden.

HUMBACH / FAISS 2010

HUMBACH, Helmut and FAISS, Klaus.

2010 *Zarathushtra and his antagonists*. Wiesbaden.

JAMISON / BRERETON 2014

JAMISON, Stephanie W. and BRERETON, Joel P.

2014 *The Rigveda. The earliest religious poetry of India* translated. I–III. (South Asia Research). Oxford – New York.

KELLENS 2006–2011; KELLENS 2013; KELLENS 2015; KELLENS 2017; KELLENS 2021–2022

KELLENS, Jean.

2006–2011 *Études avestiques et mazdéennes*. Vol. 1: *Le Ratauuō vīspe mazišta (Yasna 1.1 à 7.23, avec Visprad 1 et 2)*. (Persika 8). 2006. Vol. 2: *Le Hōm Stōd et la zone des déclarations (Y7.24 – Y5.4, avec les intercalations de Vr3 à 6)*. (Persika 10). 2007. Vol. 3: *Le long préambule du sacrifice (Yasna 16 à 27.12, avec les intercalations de Visprad 7 à 12)*. (Persika 15). 2010. Vol. 4: *L'acmé du sacrifice. Les parties récentes des Staota Yesniia (Y27.13 – Y59) avec les intercalations de Visprad 13 à 24 et la Dahmā Āfriti (Y60 – 61)*. (Persika 16). 2011. Paris.

2013 Le jour se lève à la fin de la Gāthā *ahunauuaitī*. *Journal Asiatique* 301/1: 53–84.

2015 Retour aux premiers chapitres du Yasna. *Estudios Iranios y Turanios* 2: 123–129.

2017 Reflets du début du Yasna. *Estudios Iranios y Turanios* 3: 67–72.

2021–2022 *Études avestiques et mazdéennes*. Vol. 7: *Essai sur la Gāthā spəntā.mainiiu-*. (Persika 22). 2021. Vol. 9: *Complémentarité des deux dernières Gāthās (Y51 et Y53–54.1)*. (Persika 24). 2022.

KRICK 1982

KRICK, Hertha.

1982 *Das Ritual der Feuergründung (Agyādheya)*. Wien.

LUBOTSKY, forthcoming

LUBOTSKY, Alexander.

forthcoming *Etymological Dictionary of Indo-Iranian*. Leiden – Boston.

MAYRHOFER 1973; MAYRHOFER 1977; MAYRHOFER 1979; MAYRHOFER 1987;

MAYRHOFER 1992–2001

MAYRHOFER, Manfred.

1973 *Onomastica Persepolitana. Das altiranische Namengut der Persepolis-Täfelchen*. Unter Mitarbeit von János HARMATTA, Walter HINZ, Rüdiger SCHMITT, Jutta SEIFFERT. (Sitzungsberichte der Österreichischen Akademie der Wissenschaften. Phil.-hist. Klasse 286). Wien.

1977 *Zum Namengut des Avesta*. (Sitzungsberichte der Österreichischen Akademie der Wissenschaften. Phil.-hist. Klasse 308,5). Wien.

1979 *Die altiranischen Namen*. (= *Iranisches Personennamenbuch*, Band I, Fas-

- zikel 1–3). (Österreichische Akademie der Wissenschaften. Phil.-hist. Klasse: Sonderpublikation der Iranischen Kommission [1]). Wien.
- 1987 Artatāma. *Encyclopaedia Iranica* II: 653a
- 1992–2001 *Etymologisches Wörterbuch des Altindoarischen*. (Indogermanische Bibliothek: Zweite Reihe, Wörterbücher). I: A–D. [1986–]1992. II: N–H. [1992–] 1996. III: *Die jüngere Sprache*. [1996–]2001. Heidelberg.
- MILLS 1887
- MILLS, Lawrence Heyworth.
- 1887 *The Zend-Avesta*. Volume 3: *The Yasna, Visparad, Âfrînaġân, Gâhs, and miscellaneous fragments* translated. (Sacred books of the East 31). Oxford.
- NARTEN 1986
- NARTEN, Johanna.
- 1986 *Der Yasna Haptaŋhâiti*. Wiesbaden.
- OBERLIES 2012
- OBERLIES, Thomas.
- 2012 *Der Rîgveda und seine Religion*. Berlin.
- OLDENBERG 1967; OLDENBERG 1993
- OLDENBERG, Hermann.
- 1967–1993 *Kleine Schriften*. Teil 1–2 ed. by Klaus J. JANERT. (Glasenapp-Stiftung 1). Wiesbaden 1967. Teil 3, Stuttgart 1993.
- RAULWING 2007–2008 [2009]
- RAULWING, Peter:
- 2007–2008 „Einiges zu den Skythen, ihrer Sprache, ihrem Nachleben“. Linguistische und rezeptionsgeschichtliche Bemerkungen zum gleichnamigen Buch von Manfred Mayrhofer“, *Die Sprache* 47: 48–76.
- SADOVSKI 2004; 2007; 2021a; 2012b; 2013; 2016; 2018a; 2018b; 2018c; 2019 [2020]; 2020; forthcoming
- SADOVSKI, Velizar.
- 2004 Indoiranische Komposita und freie Syntagmata mit prädikativem bzw. partitivem Attributionsverhältnis der Elemente. In: CLACKSON, James P. T. / OLSEN, Birgit Anette (eds.): *Indo-European Word Formation: Proceedings of the Conference held at the University of Copenhagen, October 20th – 22nd 2000*. (Copenhagen Studies in Indo-European 2). Copenhagen: 295–334.
- 2007 Epitheta und Götternamen im älteren Indo-Iranischen. Die hymnischen Namenkataloge im Veda und im Avesta (Stilistica Indo-Iranica, I.). In: PANAINO, Antonio / SADOVSKI, Velizar: *Disputationes Iranologicae*

- Vindobonenses, I.* (Sitzungsberichte der Österreichischen Akademie der Wissenschaften. Phil.-hist. Klasse 764: Veröffentlichungen zur Iranistik 41). Wien: 37–74 [& 75–108].
- 2012a Structure and Contents of Lists and Catalogues in Indo-Iranian Traditions of Oral Poetry. (Speech and Performance in the Veda and Avesta, II)'. In: KLEIN, Jared S. / YOSHIDA, Kazuhiko (eds.): *Indic across the millennia: from the Rigveda to modern Indo-Aryan. Proceedings of the Linguistic Section, 14th World Sanskrit Conference, Kyoto, Japan, September 1st – 5th, 2009*. Bremen: 153–192.
- 2012b „An den Namen werdet ihr sie erkennen“. Österreichisches Know-how in der Iranistischen Lexikographie. In: *Akademie aktuell. Zeitschrift der Bayerischen Akademie der Wissenschaften*, 2012/1: 30–33.
- 2013 Zur Morphologie und Semantik von Namen und Epitheta im Indo-Iranischen. In: SADOVSKI, Velizar / PANAINO, Antonio: *Disputationes Iranologicae Vindobonenses, II.* (Sitzungsberichte der ÖAW. Philosophisch-historische Klasse, 845. Band / Veröffentlichungen zur Iranistik, Nr. 65). Wien: 5–45.
- 2016 “The columns of Ṛta”: Indo-Iranian lexicon and phraseology in the ritual poetry of the Avesta, Veda and Beyond. In: HAJNAL, Ivo / KÖLLIGAN, Daniel / ZIPSER, Katharina (eds.): *Miscellanea Indogermanica. Festschrift für José Luis García Ramón zum 65. Geburtstag.* (Innsbrucker Beiträge zur Sprachwissenschaft 154). Innsbruck: 725–749.
- 2018a Ritual Formulae, Structures, and Activities in Vedic and Avestan Liturgies between Cultic Practice, Mythology, and Social Ideology. *MSS* 71/1: 95–133.
- 2018b Nominalkomposita, Neowurzelbildungen und zugrundeliegende syntaktische Konstruktionen im Veda und dem Avesta. *DABIR* 6: 156–167.
- 2018c Vedic and Avestan Parallels from Ritual Litanies and Liturgical Practices I. In: *Farnah. Indo-Iranian and Indo-European Studies in Honor of Sasha Lubotsky*. Ann Arbor: 307–326.
- 2019 [2020] Performative Speech Acts in Indo-Iranian Sacred Jurisprudence: Between Ritual Formulation, Pragmatic Application, and Political Re-semanticization. In: BADALKHAN, Sabir – BASELLO, Gian Pietro – DE CHIARA, Matteo (eds.): *Iranian Studies in Honour of Adriano V. Rossi*. Part Two. (Università degli Studi di Napoli “L’Orientale”. Dipartimento Asia, Africa e Mediterraneo. Series Minor 87,2). Naples: 797–825.
- 2020 Indo-Iranian **HamHa-* and the fighter’s ‘*Sturm und Drang*’, in compound epithets, personal names, and poetical phraseology from the Veda, Avesta and Ancient Persia. (Ritual Formulae and Ritual Pragmatics in the Veda and Avesta, IV)'. In: Repanšek, Luka – Bichlmeier, Harald – Sadovski, Velizar (eds.): *vácār̥si miśrā́ kṛṇavāmahai. Akten der Arbeitstagung der In-*

- dogermanischen Gesellschaft und des IWoBA XII vom 4. bis 7. Juni 2019 in Ljubljana aus Anlass der Hundertjahrfeier der dortigen Indogermanistik. (Studien zur Historisch-Vergleichenden Sprachwissenschaft 17). Hamburg: 593–617.
- forthcoming Knowledge of magic – magic of knowledge. In: BADALANOVA-GELLER, Florentina (ed.): *Knowledge to die for. Proceedings of the Conference held at the Max-Planck-Institut für Wissenschaftsgeschichte, Berlin 2011*. Leiden – New York etc.: 216–241.
- SCARLATA 1999
 SCARLATA, Salvatore.
 1999 *Die Wurzelkomposita im R̥g-Veda*. Wiesbaden.
- SCHINDLER 1975; SCHINDLER 1986
 SCHINDLER, Jochem.
 1975 Zum Ablaut der neutralen -s-Stämme des Indogermanischen. In: RIX, Helmut (ed.): *Flexion und Wortbildung. Akten der V. Fachtagung der Indogermanischen Gesellschaft Regensburg, 9. – 14. September 1973*. Wiesbaden: 259–267.
 1986 Zu den homerischen ῥοδοδάκτυλος-Komposita. In: ETTER, Annemarie (ed.): *ooperosi. Festschrift für Ernst Risch zum 75. Geburtstag*. Berlin – New York: 393–401.
- SCHMITT 2003; SCHMITT 2011; SCHMITT 2017
 SCHMITT, Rüdiger.
 2003 Onomastische Bemerkungen zu der Namenliste des *Fravardin Yašt*. In: CERETI, Carlo G. / MAGGI, Mauro / PROVASI, Elio (eds.): *Religious themes and texts of pre-Islamic Iran and Central Asia*. Wiesbaden: 363–374.
 2011 *Iranische Personennamen in der griechischen Literatur vor Alexander d. Gr. = Iranisches Personennamenbuch*, Band V, Faszikel 5A. (Sitzungsberichte der ÖAW. Phil.-hist. Klasse, 823: Iranische Onomastik 9). Wien.
 2017 The Lexicon of Indo-Iranian. In: KLEIN, Jared S. / JOSEPH, Brian / FRITZ, Matthias (eds.): *Handbook of Comparative and Historical Indo-European Linguistics*. In cooperation with Marc WENTHE. (Handbücher zur Sprach- und Kommunikationswissenschaft 41,3). Berlin – New York: 1942–1959.
- SCHULZE 1892
 SCHULZE, Guilelmus [Wilhelm].
 1892 *Quaestiones epicae*. Gueterslohiae [Gütersloh].
- SCHUTZEICHEL 2013
 SCHUTZEICHEL, Marc.
 2013 Indogermanische Funktionsverbgefüge. Diss., Univ. Münster. Münster.

SWENNEN 2015

SWENNEN, Philippe.

- 2015 Indo-iranien **niūajdaja-* : le mécanisme de l'annonce liturgique. *Estudios Iranios y Turanios* 2: 209–217.

TAVERNIER 2007

TAVERNIER, Jan.

- 2007 *Iranica in the Achaemenid Period (ca. 550–330 B.C.). Lexicon of Old Iranian Proper Names and Loanwords, Attested in Non-Iranian Texts.* (Orientalia Lovaniensia Analecta 158). Leuven – Paris – Dudley, MA.

WEBER 1849

WEBER, Albrecht (ed.).

- 1849 *The White Yajurveda. Vol. 1: The Vâjasaneyi-Sanhitâ.* Berlin.

WEISS 2009

WEISS, Michael.

- 2009 Umbrian **erus**. In: YOSHIDA, Kazuhiko (ed.): *East and West: Papers in Indo-European studies.* Bremen: 241–264.

ZADOK 2009

ZADOK, Ran.

- 2009 *Iranische Personennamen in der neu- und spätbabylonischen Nebenüberlieferung = Iranisches Personennamenbuch, Band VII, Faszikel 1B.* (Sitzungsberichte der ÖAW. Phil.-hist. Klasse, 777 / Iranische Onomastik 4). Wien.

ARTICOLI, NOTE, SAGGI

*Analisi linguistiche di testi arcaici, riflessioni su aspetti e problemi linguistici
del mondo antico, linee e momenti di preistoria e protostoria linguistica*

CRISTINA RESMINI*

Grammaticalisation paths of the Proto-Indo-European anaphoric pronoun **tód* across the Germanic languages

Abstract

Secondo l'etimologia generalmente riconosciuta e nel contesto della grammaticalizzazione, si ritiene che, nel corso del tempo, i riflessi germanici del pronome anaforico protoindoeuropeo **tód* abbiano assunto le funzioni di pronome dimostrativo, determinatore, pronome relativo, complementatore argomentale e congiunzione consecutiva. Sulla base dei dati empirici raccolti attraverso un corpus linguistico creato *ad hoc* relativo alle fasi più antiche delle lingue germaniche e alla luce di evidenze cross-linguistiche provenienti specialmente dal latino, questo articolo propone che i riflessi germanici di **tód*, che inizialmente fungevano da pronomi dimostrativi, siano diventati determinatori laddove seguiti da un sostantivo e siano stati poi direttamente rianalizzati in complementatori argomentali attraverso la struttura del dittico inverso. L'argomentale, a sua volta, sarebbe stato successivamente rianalizzato per estensione in congiunzione consecutiva in contesti consecutivi e in pronome relativo improprio in presenza di antecedente.

Parole chiave: lingue germaniche, grammaticalizzazione, pronome anaforico protoindoeuropeo, pronome relativo, complementatore argomentale

Based on the accepted etymology and within a grammaticalisation framework, the Germanic reflexes of the Proto-Indo-European anaphoric pronoun **tód* are believed to have diachronically acquired the functions of demonstrative pronoun, determiner, relative pronoun, argumental complementiser, and consecutive conjunction. According to the empirical data gathered from an extensive Early Germanic linguistic corpus created specifically for this study and considering cross-linguistic evidence especially from Latin, this paper suggests that the Germanic reflexes of **tód* – which initially functioned as demonstrative pronouns – became determiners when followed by a noun and were directly reanalysed into argumental complementisers through a *dipthique inverse* construction. In turn, the argumental complementiser was then reanalysed by extension into a consecutive conjunction in consecutive contexts, and into a relative *d-Pronomen* in the presence of an antecedent.

Key words: Germanic languages, grammaticalisation, proto-indo-european anaphoric pronoun, relative pronoun, argumental complementiser

* Cristina Resmini, Università di Napoli L'Orientale, c.resmini@unior.it.

1. Introduction

This article intends to investigate the grammaticalisation processes which have led the Present-Day Germanic reflexes of the Proto-Indo-European anaphoric pronoun **tód* to acquire multiple functions (i.e. demonstrative pronoun, determiner, relative pronoun, argumental complementiser and consecutive conjunction).

To this purpose, this research traces the evolution of the reflexes of **tód* throughout the earliest stages of the Germanic languages by means of a linguistic corpus which was created specifically for this study and which includes attestations from all three Germanic branches, dating from the second to the fourteenth century CE.

Since the corpus was meant to serve as the basis for a comparative analysis of the three Germanic branches, the guiding principle in building it was to ensure that each branch would be equally represented. For this reason, the corpus mainly focuses on one single language per branch, i.e. Gothic for the East Germanic branch, Old Norse for the North Germanic branch, and Old English for the West Germanic one. While in the first two cases the selected languages are the only possible option, the choice to examine Old English, instead of any of the other West Germanic languages, is deliberate and it is due to several reasons. Firstly, the Old English written production is much greater than that of any other West Germanic language and it thus represents the only West Germanic production that is comparable in size to that of Gothic and Old Norse. Secondly, most studies in the field seem to have often neglected Old English in favour of Old High German – where some relevant structures are undoubtedly more apparent. However, Old English presents us indeed with interesting dilemmas, as there exist contrasting accounts on a possible early grammaticalisation of relatives – which is exactly the crux of this study. Nevertheless, the linguistic corpus prepared for this research includes a number of relevant entries from other West Germanic languages (i.e. Old High German, Old Low Franconian, Old Frisian and Old Saxon), which appear to be consistent with what can be observed in Old English.

According to the accepted etymology and within the frame of a grammaticalisation process, the Early Germanic reflex of **tód* – which was originally a demonstrative pronoun – become a determiner when followed by a noun and

was then supposedly reanalysed into a relative pronoun through the *dipthique normal* structure (Viti 2013). In the presence of verbs such as *verba dicendi* and *sentiendi*, the relative thus assumedly grammaticalised by extension into an argumental complementiser (Axel-Tober 2017), through the *dipthique inverse* structure, and then, by over-extension, into a consecutive conjunction, when found within a consecutive context.

While, as far as each individual language is concerned, this study's findings are indeed representative and consistent with the existing literature, the above-mentioned model in its entirety does not, however, apply to the empirical data gathered from the corpus. In fact, based on those data, the argumental and consecutive functions appear to have grammaticalised way before the relative one.

In order to explain such discrepancy, we must thus resort to cross-linguistic evidence (Narrog and Heine 2018). By comparing the Latin correlative structures (De Roberto 2010) with the Germanic ones, we could, in fact, argue that the *dipthique normal* seems to have only given origin to the Germanic *w-Pronomen*. Following on this hypothesis, this paper then suggests that, on the one hand, the argumental complementiser was grammaticalised directly from the demonstrative, through a *dipthique inverse* construction, and that, on the other hand, consecutive conjunctions and relative *d-Pronomen* derive from a later extension of the meaning of the argumental complementiser itself.

While the initial sections of this article present a detailed literature review [cf. 2. Theoretical Background] and a description of the linguistic corpus that has been compiled for this study [cf. 3. Methodology], the core of this paper is dedicated to the analysis of the data gathered from the corpus [cf. 4. Corpus Analysis] and to the proposition of a model which, based on that empirical evidence, aims to account for the distinct origins of argumental complementisers and relative pronouns [cf. 5 Discussion]. Finally, following a short summary of the main points this paper tries to make [cf. 6. Conclusion], a selection of the relevant corpus entries mentioned throughout the article is included in the Appendix section.

2. Theoretical Background

In Present-Day English, the word *that* covers multiple functions – including those of demonstrative and relative pronoun, determiner, and conjunction –

and the same can be said, for instance, of *das(s)*¹ and *dat* in, respectively, Present-Day German and Dutch. For this reason, in order to analyse the grammaticalisation paths which have led these morphemes to acquire all of the above-mentioned functions, it is necessary to provide an overview of the way in which the main Present-Day Germanic languages (i.e. English, German, Dutch, Danish, Norwegian, Swedish, and Icelandic) express those functions.

	EN	DE	NL	DA	NO	SV	IS
Demonstrative Pronouns/ Determiners (Proximal, Distal)	<i>this</i> , <i>that</i>	<i>der-die-das/dieser-diese-dieses/derjenige-diejenige-dasjenige, der-die-das dort/jener-jene-jenes</i>	<i>deze-dit, die-dat</i>	<i>den-det</i>	<i>den-det, denne-dette</i>	<i>den-det här, den-det där</i>	<i>þessi-þetta, sá-sú-það</i>
Definite Articles	<i>the</i>	<i>der-die-das</i>	<i>de-het</i>	<i>-en/-et, den-det</i>	<i>-en/-a/-et</i>	<i>den-det, -en/-et</i>	<i>-inn/-n/-ið, hinn</i>
Relative Pronouns (<i>w-Pronomen, d-Pronomen</i>)	<i>who-what, that</i>	<i>wie-was, der-die-das</i>	<i>wie-wat, die-dat</i>	<i>hvo-hvad, som, der</i>	<i>som</i>	<i>som</i>	<i>sem, er</i>
Conjunctions (ARG, CONS)	<i>(that)</i> , <i>so that</i>	<i>(dass)</i> , <i>sodass/ damit</i>	<i>(dat)</i> , <i>omdat/ zodat</i>	<i>(at)</i> , <i>så at</i>	<i>(at)</i> , <i>Slik at/ Så at</i>	<i>(att)</i> , <i>Så att</i>	<i>(að)</i> , <i>svo að</i>

Tabella 1. Forms per function across the main Germanic languages.

2.1. From pronoun to article

Historical linguists and etymologists alike agree on the fact that the equivalents of *that* (i.e. *das(s)*, *dat*, *det*, and *það* – which are indicated **in bold** in Table 1) are indeed its cognates and that their common ancestor is the neuter form of the Proto-Indo-European anaphoric pronoun **só-*, **sá-*, **tód* (Sijts 2010).

PIE **tód* > PG **þat* > OE *thæt*, OHG *thaz/daz*, OLF *that*, ON *þat* >
EN *that*, DE *das*, NL *dat*, DA *det*, NO *det*, SV *det*, IS *það*

¹ The different distribution of *das* and *dass* is due to a mere spelling difference, probably introduced to avoid ambiguity (Axel-Tober 2017).

According to Quiles and López-Menchero (2011), the Proto-Indo-European *só-, *sá-, *tód was an anaphoric demonstrative pronoun which tended to appear at the beginning of sentences, as it had probably originated from a relative pronoun. It expressed a *that*-deixis, thus referring to something that was generally *there*, without any specific spatial reference, and could be translated as either *this* or *that*.

As it is evident from Table 1, *that* and its cognates do indeed still act as demonstrative pronouns in all of the Present-Day Germanic languages taken into account and this indicates that the original deictic meaning of *tód has been retained to this day, as well as its pronominal and anaphoric nature. Since it is therefore clear that the Present-Day Germanic demonstrative pronoun is the immediate reflex of the original *tód, it seems acceptable to take this specific function as the starting point of this analysis.

Today, though, the Germanic demonstrative pronouns may also be used adjectivally as demonstrative determiners. At the time when Proto-Germanic split from Proto-Indo-European, in fact, the articles system had not yet been introduced in the mother language and, when the difference between indefinite and definite noun was considered to be crucial, the anaphoric pronoun *só-, *sá-, *tód could also act as a definiteness marker. At that stage of Proto-Indo-European, however, the relationship between the demonstrative and the name to which it referred is assumed to have been of an appositional nature (Quiles and López-Menchero 2011), rather than of an adjectival one – which would otherwise imply that the category of determiners was already in existence. Indeed, it was only with the diachronic emergence of definite articles and the resulting appearance of the category of definiteness that a change in the syntactic structure of the languages occurred and a determiner phrase projection was created (Lyons 1999).

With the emergence of the category of determiners, the Germanic demonstrative pronoun acquired the function of demonstrative determiner, while its role as a definiteness marker was then transferred to the rising definite article, with fairly heterogeneous outcomes among the various languages. On the one hand, German retains the morpheme *das* also for the neuter article, while the English *the* and the Dutch *de* both just represent a different evolution of the original *tód (Sijts 2010). On the other hand, Norwegian,

Icelandic, Faroese, Danish, and Swedish articles are generally expressed by means of suffixes – although the latter two languages also use the morpheme *det* as the neuter form of the full article, either as an alternative to (Danish) or together with (Swedish) the definite suffix.

The grammaticalisation path which led the anaphoric demonstrative pronoun to become a definite article is considered to be an “innovation,” in that an entirely new grammatical category has emerged from this process (Lehmann 2002). At an initial stage, the demonstrative could be used freely (i.e. without context constraints) as an apposition of the head noun by which it was followed and had a [+deictic] feature. However, at some point, this construction was extended to contexts where the deictic reference was no longer relevant – i.e. when the demonstrative was used to indicate that its head noun had already been mentioned. In this specific context, where an anaphoric reference is made, the foregrounded [+anaphoric] feature of the demonstrative acted as the bridging context which brought about an extension of the meaning of the pronoun itself, through a context-induced reinterpretation. Since this new context was incompatible with the original deictic meaning of the demonstrative, though, the latter lost its [-deictic] function and gave rise to the innovated feature of [+definiteness], which is indeed based on anaphoric reference. It is thus during the final phase of the process that the new meaning was conventionalised and the demonstrative could therefore be used freely with that new acceptance, even in other contexts. As already mentioned, some Germanic languages have retained the same word both for the article and for the demonstrative, while, probably to avoid ambiguity, some others – like English and Dutch – have adopted a different and decategorised version of the demonstrative as their definite article. Yet, other languages (i.e. Norwegian, Icelandic, and Faroese) do not seem to have undergone this process at all [cf. Section 4.2].

2.2. From pronoun to subordinator

As far as the grammaticalisation of the anaphoric pronoun is concerned, however, the origin of definite articles is only the first of several differences between North and West Germanic languages. As shown in Table 1, in fact, while in English, German, and Dutch *that*, *das*, and *dat* may also be used –

among other forms – as relatives, in no Scandinavian language do the demonstratives *det* and *það* cover this specific function. This indicates the existence of two different isoglosses as far as the use of that and its cognates is concerned, but also raises questions about what might have caused the two sub-families to take these two diverging paths.

In the North Germanic languages, the relative function is generally expressed by relative particles [cf. Section 4.3.3], with the sole exception of Danish, which also features the relative pronoun *hvo*, *hvad*. This pronoun, is a cognate of the English *who*, *what*, the German *wer*, *was* and the Dutch *wie*, *wat*, and they all are reflexes of the original Proto-Germanic adjective interrogative-indefinite pronoun **qos*, **qā*, **qod* (Proto-Indo-European stem **k^wó-*) – which, in turn, had probably originated from an even more ancient Proto-Indo-European relative pronoun (Quiles and López-Menchero 2011).

PIE **k^wó-* > PG **h^wa-* >

EN *who*, *what*; DE *wer*, *wie*; NL *wie*, *wat*; DA *hvo*, *hvad*

Fuß and Grewendorf (2014) refer to these Germanic reflexes as *w-Pronomen*, which usually head free relative clauses and are considered as the only “proper” relative pronouns. On the other hand, they also stress the existence, in the West Germanic languages, of the so-called *d-Pronomen*, which come from the surface-identical demonstrative and usually introduce headed relative clauses [cf. Section 5]. At a closer look, however, the English relative *that* can hardly be considered as a fully fledged pronoun [cf. Section 4.3.2]. Its invariable form, in fact, could lead to it being interpreted as a relative particle (and thus a relative complementiser), even though, from a diachronic perspective, it could also be seen as a “degenerate resumptive pronoun,” i.e. a pronoun which has gradually lost its inflectional properties (Haspelmath 2001: 1494-1495).

From this brief account of the Present-Day Germanic relatives system, it is therefore clear that – at least as far as the West Germanic languages are concerned – there are two competing patterns of development in the formation of relatives, where the relative *w-Pronomen* derive from the interrogative-indefinite pronoun and the *d-Pronomen* originate from the anaphoric demonstrative one. For the purposes of this study, however, the most relevant forms are indeed those which result to be identical to the demonstrative.

The intent is, in fact, that of analysing if and how a demonstrative could be grammaticalised into a relative, especially because this pattern seems to be quite specific to this particular sub-group of languages.

Although, traditionally, Proto-Indo-European was assumed to lack subordinate structures and the hypotactic constructions in its daughter languages were believed to have descended from paratactic ones, most scholars nowadays agree on ascribing “at least relative clauses and some types of adverbial clauses” to Proto-Indo-European itself (Viti 2013: 89). The earliest reconstructed complex structure is, however, not yet a form of subordination, but a form of correlation, i.e. the correlative diptych identified by Minard (1936) in Vedic and later found by other linguists in other Indo-European languages, as well. Correlation is a relation between two clauses in which none of the two can exist independently from the other. While this makes it a relation that is unlike either coordination or subordination, Haudry (1979) maintains it might still be considered as the form of relation from which subordination stemmed. The correlative diptych is, in fact, neither an independent, nor an embedded structure (yet), but rather “a subtype of adjoined relative clause” (Hendery 2012: 18), where – in the case of a *dyptique normal* (Minard 1936) – “a preposed relative clause presents a noun phrase, which is anaphorically resumed in the subsequent main clause by a demonstrative pronoun” (Viti 2013: 92).

- (1) a. DE *Was er macht, das mach-t er richtig.
 REL.ACC.N.SG he does DEM.ACC.N.SG does he right.ADV²
 ‘What(ever) he does, he does that right’
- b. DE Was er mach-t, mach-t er richtig.
 REL.ACC.N.SG he does does he right.ADV
 ‘What(ever) he does, he does right’

In particular, as shown in (1a), the *dyptique normal* consists in the adjoining of two simple – but not independent – clauses, the first of which is introduced by an element which belonged to the interrogative-indefinite pronoun-sys-

² All the glosses throughout the paper follow the Leipzig glossing rules and the information provided is generally restricted to what is specifically relevant for the scopes of this study.

tem (e.g. *was*), while the second one is introduced by an anaphoric demonstrative pronoun (e.g. *das*). However, over time, one of the two elements has been absorbed by the other, thus resulting in sentence (1b). The merger of the two elements – or the omission of one of the two – is what has historically led the correlative diptych to become a hypotactic structure, where the clause “*was er macht*” is embedded in the main clause (i.e. “*macht er richtig*”), as well as what has allowed for the originally interrogative-indefinite pronoun to become a proper relative pronoun.

Although this example manages to show the origin of relative subordinates, it does not yet explain if or how the demonstrative could ever become a relative – and the fact that the demonstrative pronoun is dropped would rather seem to imply it cannot. Let us thus consider the following examples:

- (2) a. EN *I have read the book **that** you gave me.*
 b. DE *Ich habe das Buch gelesen, **das** du mir gegeben hast.*
 I have ART.ACC.N.SG book read.PP DEM.ACC.N.SG you to.me given have
 ‘I have read the book that you have given me’
 c. NL *Ik heb het boek gelezen **dat** je me had gegeven.*
 I have ART.N.SG book read.PP DEM.N.SG you to.me have given.PP
 ‘I have read the book that you had given me’

These three sentences instantiate the use of *that*, *das*, and *dat* as relatives in Present-Day West Germanic languages. In order for those words to begin to be used as such, however, we have to assume a precedent correlative structure as in (3a), (3b) and (3c):

- (3) a. EN **I have read the book **that** which you gave me.*
 b. DE **Ich habe das Buch gelesen, **das** was*
 I have ART.ACC.N.SG book read.PP DEM.ACC.N.SG REL.ACC.N.SG
 du mir gegeben hast.
 you to.me given.PP have
 *‘I have read the book that which you have given me’
 c. NL **Ik heb het boek gelezen **dat** wat je me*
 I have ART.N.SG book read.PP DEM.N.SG REL.N.SG you to.me
 had gegeven.
 have given.PP
 *‘I have read the book that which you had given me’

In these cases, as the Present-Day outcomes show (2a), (2b), and (2c), the resumptive pronoun has prevailed over the proper relative pronoun – which has consequently been dropped – and the anaphoric demonstrative pronoun has become a relative pronoun in its own right.

Nevertheless, since English, German, and Dutch all retain both the *w-Pronomen* and the *d-Pronomen* forms of the relative we must consider this grammaticalisation process as a “renovation” (Lehmann 2002) that is only partly innovative, as indeed the new relative (the *d-Pronomen*) has never completely replaced the original proper relative. Furthermore, while in German and Dutch there is a complementary distribution among the two pronouns, in English, the relative complementiser *that* can always be replaced by a *w-Pronomen*, meaning that it is in an overlapping distribution with pronouns like *who* and *which* (Zimmermann 2012).

Beside the relative, however, there are two more grammaticalisation processes among the West Germanic languages in which the words *that*, *das*, and *dat* have been involved and these are the paths which have allowed for them to also acquire the functions of argumental complementiser and consecutive conjunction.

In historical linguistics studies, the origin and development of the argumental clause is far from being uncontroversial. From the Neogrammarians period and up to few years ago, the argumental complementiser in the West Germanic languages was traditionally believed to have derived directly from the Proto-Indo-European anaphoric pronoun **tód*. The anaphoric pronoun was, in fact, assumed to have been used cataphorically, when found within a paratactic structure and in the presence of *verba dicendi* and *verba sentiendi*.

- (4) a. EN I know that you are wrong.
 b. EN *I know this: **that** you are wrong.

In example (4a), for instance, the presumed positioning of the clause boundary would be between *that* and *you* (i.e. “I know that | you are wrong”), and the demonstrative would thus indeed constitute a cataphoric reference to the following clause (i.e. “you are wrong”). Therefore, based on this traditional account, the process of reanalysis of the demonstrative into an argumental complementiser would consist in the cataphoric demonstrative moving from

the end of the first clause to the beginning of the second one, through a shift of clause boundary, and in the paratactic structure consequently developing into a hypotactic – albeit surface-identical – construction.

In recent years, however, this hypothesis has been abandoned in favour of a new one which sees the argumental complementiser as originating from the relative, instead. As most subordinators, in fact, the argumental clause is cross-linguistically based on the relative one (Haudry 1979) and it seems only fitting that the complementiser introducing it should have originated from some form of relative pronoun or complementiser. As further confirmation of that, it is believed that the argumental clause had derived from a correlative structure, i.e. the *diptyque inverse* (Minard 1936), where – contrarily to the *diptyque normal* – “the head noun is contained in the main clause, which precedes the relative clause” (Viti 2013: 92), or, in this case, the argumental one. This correlative structure thus consists of two adjoined clauses: the first one presents a demonstrative pronoun used cataphorically, while the second one is introduced by a relative (4b) – which can either be a relative complementiser as in English, or the neuter form of the relative pronoun, as in German and Dutch.

According to this second hypothesis, which has been especially put forward by Axel-Tober (2017), the argumental complementiser is indeed assumed to derive from a correlative structure where one element is the optionally silent demonstrative and the other element is the already grammaticalised relative *d-Pronomen*. In this case, the demonstrative could easily be dropped in order to simplify the structure and the *d-Pronomen* would thus prevail with no shift of clause boundary because it has always belonged to that second clause.

Therefore, we may maintain that, when in the presence of some specific categories of verbs [cf. Section 4.4] the relative *d-Pronomen* has been reanalysed as an argumental complementiser, through a context-induced “extension” of its meaning.

On the other hand, the grammaticalisation path which has seen *that* and its West Germanic cognates take on the function of consecutive conjunction is an “overextension,” in that it represents the extension in meaning of an already extended element, such as the argumental complementiser. When in

the specific context of the consecutive construction (e.g. *so+ADJ/ADV...that*), in fact, the argumental complementiser is assumed to have undergone a “semantic bleaching” in which it has lost its semantic function and acquired a merely syntactic one.

As for the argumental complementiser and consecutive conjunction in the North Germanic languages, the forms reported in Table 1 are *at* (Danish, Norwegian), *att* (Swedish), and *að* (Icelandic). Although these morphemes are formally different from the Scandinavian demonstratives and even from the West Germanic *that*, *das*, and *dat*, a reconstruction which sees them as just different reflexes of the same Proto-Indo-European pronoun **tód* is not only plausible, but quite popular among scholars (Harbert 2007; Bisang and Malchukov 2020).

PIE **tód* > PG **þat* > ON *þat/at* >
 Demonstratives: DA *det*, NO *det*, SV *det*, IS *það* / Conjunctions:
 DA *at*, NO *at*, SV *att*, IS *að*

Since, in fact, we would only need to postulate the dropping of the first consonant (Vigfusson 1874:28-29) for this development to be acceptable, we cannot exclude this possibility just on grounds of reconstruction. In this scenario, *at*, *att*, and *að* are cognates of *that*, *dass*, and *dat* and, although they seem to have mostly skipped the other grammaticalisation processes (i.e. demonstrative > definite article; demonstrative > relative³), they have been reanalysed into argumental complementiser and consecutive conjunction [cf. Section 4.4.3]⁴.

On the flip side, though, there seems to exist the possibility that *at*, *att*, and *að* all derive from the above-mentioned homographic preposition and conjunction, which, according to Vigfusson (1874:25-28), is a cognate of the

³ Interestingly, though, while this is no longer true in Present-Day Icelandic, Vigfusson (1874: 29) mentions that the now archaic *at* (now replaced by *að*) could also be used as a less frequent alternative of the relative particle *er*, thus implying that the grammaticalisation of the demonstrative into a relative might indeed have happened, although no trace of it is left in Present-Day Icelandic.

⁴ If this hypothesis were true, we would also have to assume a spelling difference between the demonstrative and the conjunction (not unlike that of the German *das/dass*), and to accept the existence in Present-Day North Germanic languages of a homographic preposition and conjunction.

English preposition *at* and has originated from the Proto-Indo-European preposition *ad*.

PIE **ad* > PG **at* > ON *at* > DA *at*, NO *at*, SV *att*, IS *að*

The preposition *ad*, which meant *to*, *by* and *near*, was then reanalysed as a subordinating conjunction (translatable as the English conjunction *to*) and, if this second hypothesis could hold, further extended to the functions of argumental complementiser and consecutive conjunction.

In the first case, North and West Germanic languages would all belong to a single isogloss, as far as most of the grammaticalisation paths of the Proto-Indo-European **tód* are concerned. In the second case, on the other hand, we would be in the presence of two very distinct isoglosses among sister languages featuring cognate demonstratives. Although traditionally the first hypothesis seems to be the most accredited one, it is still interesting to analyse the validity of the second one, as it would appear to be plausible, not only from a linguistic point of view, but also from a historical one. Indeed, the discriminating factor in the matter is presumed to be the influence of Latin syntax on the Germanic languages, which has been of a major significance among the West Germanic languages, but virtually non-existent on the North Germanic ones.

2.3. Grammaticalisation and typology

This concept, however, raises the long-standing question of what could be the triggering factors which may lead to the grammaticalisation of some specific elements of a language. According to Heine and Kuteva (2010:380), grammaticalisation is the unidirectional process which allows for lexical forms to develop into grammatical forms, and for grammatical forms to evolve into even more grammatical ones. Assuming that every kind of linguistic evolution is the result of language change, when it comes to grammaticalisation, we have to consider that it is mainly motivated by the need to use “linguistic forms that are concrete, easily accessible, and/or clearly delineated to also express less concrete, less easily accessible, and less clearly delineated meaning contents” (Heine and Kuteva 2010: 381). In order for these processes to be set into motion, though, we need to also factor in a certain amount of creativity, “whereby

existing material is combined in novel ways to produce new artefacts” – i.e. new linguistic meaning and structures (Heine and Kuteva 2010: 381).

There are several accounts on how grammaticalisation processes may occur within a language, but most scholars have favoured either a transformational-generative approach or a typological perspective as the theoretical framework for their studies.

The transformational-generative approach is especially grounded in Chomsky’s notions of “deep structure” and “surface structures” as the two coexisting levels of representation within a language. The first one contains all the information relevant to semantic interpretation, while the latter contains all the information relevant to phonetic interpretation alone (Chomsky 1988: 14-15). According to this approach, syntactic change is generally associated with child language acquisition, in the sense that it is seen as part of what happens in the transition of grammars from one generation to the next⁵ (Campbell 1998:234-235). The child learners, in fact, are presumed to construct their own grammar on the basis of the input they receive from the adults around them. However, learners’ acquisition of grammar will be limited to the surface structure of the input they have received and processed, and there is thus no guarantee that the grammar they have acquired does coincide with the deep structure of the adults’ grammar. Furthermore, language speakers may at some point add rules to their grammar which would make it no longer optimal⁶ and, it will thus be up to the children of the following generation to restructure the non-optimal adult grammar and make it more optimal, as they construct their own internal grammar. Indeed, Kroch (2001: 2) argues that language change is “by definition a failure in the transmission across time of linguistic features,” while Roberts and Roussou (2007) claim that syntactic change can be interpreted as parametric variation. Finally, in more recent works, Biberauer & Roberts (2017: 134) attempt to integrate “the concerns of historical syntax with those of generative grammar” and

⁵ According to Heine and Kuteva (2010: 382), “a *sine qua non* for grammaticalisation is the transmission of linguistic knowledge from one generation to the next.”

⁶ Optimality occurs when natural language grammars create structures which are designed to interface perfectly with other components of the mind, i.e. speech and thought systems (Radford 2004: 9).

propose that the parameters of Universal Grammar are not pre-specified, as they, instead, supposedly emerge from the interaction subconsciously operated by the acquirer between the three factors of language design, i.e. the Universal Grammar, the primary linguistic data and the non-domain-specific cognitive optimisation principles.

Although such descriptions of how syntactic change occurs are certainly valid ones, they cannot provide a satisfactory answer as to why certain elements of a specific language do indeed grammaticalise whereas others do not, nor as to how unrelated languages may take similar paths, while related ones behave differently.

It thus appears evident that there must be some other factor at play, besides the ones we have already mentioned, and that is language typology. The diachronically typological approach is in fact the one that has been adopted for this study.

Typology is “the study of structural differences and similarities between languages” (Velupillai 2012: 15) and, while synchronic typology aims at listing and categorising all possible patterns, diachronic typology attempts to explain why such patterns exist and why they change (Pat-El 2021). According to Bybee (2006), some universal mechanisms lead, in fact, to pathways of change which then reflect on synchronic linguistic structures and, since these patterns are often cross-linguistically similar, their synchronic outcomes may indeed bear some resemblance. In particular, the typological features of a language are expected to correlate with some aspects of grammaticalisation, in that “certain criteria of grammaticalisation may apply differently to different types of language” and this correlation does not appear to be unidirectional, as it is argued both that “typological features influence aspects of grammaticalisation” and that “grammaticalisation motivates structural features that can be typologised” (Narrog and Heine 2018: 1-2). Therefore, whereas generative approaches tend to reduce grammaticalisation “to an essence that is universal and not amenable to typological influences” (Narrog and Heine 2018: 4), from a typological perspective, it is argued that “the likelihood for a certain grammaticalisation process to appear is at least to some extent dependent on structural properties of the language” (Dahl 2018: 95). Furthermore, Narrog and Heine (2018: 14-15) main-

tain that there exists a general tendency in languages to “follow already trodden grammaticalisation paths and to reproduce or flesh out established grammatical categories [...] rather than to create entirely new structures and categories,” while they also claim that when languages diverge from this “tendency towards conservatism, or inertia,” it is most likely due to intense language contact.

When it comes to similarities in development between languages, however, typology, tends to reject the hypothesis of language contact and prefers to explain them as parallel independent developments due to “similarities in preconditions – either internal, i.e. shared structural properties, or external – shared ecologies, or universal cognitive properties” (Dahl 2018: 79).

As descendents of a common ancestor, in fact, the Germanic languages are expected to bear varying degrees of resemblance to all of the other Indo-European languages, even though they belong to different branches and have not had direct or significant contact with them after they split from the mother language. Moreover, towards the end of the twentieth century, some scholars identified an European linguistic area, i.e. the so-called “Standard Average European”, or SAE, which mainly consists of the Romance, Germanic, Balto-Slavic, and Balkan languages, as well as the non-Indo-European westernmost Finno-Ugrian languages. As a linguistic area, the Standard Average European comprises a number of geographically contiguous languages that “share structural features which cannot be due to retention from a common proto-language and which give these languages a profile that makes them stand out among the surrounding languages” (Haspelmath 2001: 1492).

Following on these concepts, the shared properties within either a language family or a language area seem to have a major role also as far as grammaticalisation is concerned. For this reason, it is necessary to broaden the scopes of this study and further analyse the already discussed grammaticalisation paths, from an Indo-European perspective.

As mentioned earlier, articles are typically considered as a late Proto-Indo-European feature which only appeared after the daughter languages had already started to separate from the mother language. This explains why Latin and the early Homeric Greek lacked definite articles, while later forms of

Ancient Greek, the Romance languages, and even the Present-Day Germanic languages do not. However, these languages belong to different Indo-European branches and they have introduced the articles system at very different times. Interestingly, however, the grammaticalisation path which has brought them to acquire the definite article is exactly the same as the one we have already described and started with the language-specific reflex of one of the Proto-Indo-European demonstratives.

PIE **tód* > GR *tó* (ART.NOM.N.SG)
LAT *ille* (DEM.NOM.M.SG) > IT *il* (ART.M.SG)

Besides **só-*, **sá-*, **tód*, in fact, Proto-Indo-European also featured a second anaphoric demonstrative pronoun, i.e. **is*, **ieh₂-*, **id* – from which, for instance, the Latin pronoun *is*, *ea*, *id* originated and which, in turn, had derived from an even older demonstrative of stem **i-* (Quiles and López-Menchero 2011). However, according to Quiles and López-Menchero (2011), in Proto-Indo-European, there was another pronoun which had stemmed from the anaphoric stem **i-*, i.e. the original relative pronoun **ios*, **iā*, **iōd*, from which many Indo-European languages, including Indo-Iranian, Greek, Phrygian and Slavic, have derived their relative pronoun. While **ios*, **iā*, **iōd* usually introduced appositive-explicative relative clauses, however, there also existed a second relative pronoun, which introduced attributive-restrictive clauses and came from the Proto-Indo-European interrogative-indefinite pronouns **qis*, **qid* (substantive form) and **qos*, **qā*, **qod* (adjective form) – from which, as already mentioned, the Present-Day Germanic *w-Pronomen* derive. From **qis*, **qid*, on the other hand, the Latin *quis*, *quid* originated, which then produced reflexes in the Romance languages, such as the Italian *chi*, *che* (*cosa*) and the French *qui*, *que*.

As we can thus see from this account, Proto-Indo-European itself is assumed to have featured a complementary distribution between two relative pronouns which had originated from competing patterns, i.e. one from a demonstrative stem and the other from an interrogative-indefinite one. Although in Present-Day SAE languages, relatives based on a demonstrative are not common (Haspelmath 2001:1494), from a typological perspective, we may justify the fact that (West) Germanic demonstratives are assumed

to have been reanalysed into relatives, on grounds of the apparently similar origin of one of the Proto-Indo-European relatives.

As for the formation of relative, argumental, and consecutive clauses, the plausibility of the grammaticalisation paths which we have described for the (West) Germanic languages is corroborated by cross-linguistic evidence. In fact, complex sentences in Latin are believed to have originated from correlative structures, where the merger of two elements into one gave rise to different types of subordinate clauses and the merged element became a subordinator.

More specifically, according to De Roberto (2010), relative clauses in Latin derived from the *diptyque normal* structure, where the pronoun *is* had an anaphoric function and was, over time, absorbed by the interrogative-indefinite pronoun *qui*, which was thus grammaticalised into a relative pronoun (5).

- (5) LAT **Qui** *bene amat, (is)* *bene castigat.*
 REL.NOM.M.SG well loves PERS.NOM.3SG well chastises
 ‘Who well loves, he well chastises’

(De Roberto 2010)

Following on this explanation, when in the presence of *verba dicendi* and *verba sentiendi*, the already grammaticalised relative pronoun was also reanalysed into an argumental complementiser through the *diptyque inverse* structure (De Roberto 2010), where the demonstrative element was used cataphorically and, once again, dropped and absorbed by the relative (6). Moreover, just as in the (West) Germanic languages, the grammaticalised argumental complementiser was reanalysed into a consecutive conjunction, by “overextension.”

- (6) LAT **Qui** *bene amat, (is)* *bene castigat.*
 REL.NOM.M.SG well loves PERS.NOM.3SG well chastises
 ‘Who well loves, he well chastises’

(De Roberto 2010)

- (6) LAT *Legati renuntiaverunt hoc, **quod** Pompeium in*
 envoys reported DEM.ACC.N.SG REL.ACC.N.SG Pompeius in
potestate haberent.
 power had.3PL
 ‘The envoys reported this: that they had Pompeius in their power’

(De Roberto 2010)

Indeed, it thus appears that the only difference between the Germanic and the Italic grammaticalisation of relative pronouns, argumental complementisers and consecutive conjunctions, is the element they have selected to undergo those paths – in the first case, the demonstrative prevailed, whereas, in the second one, the interrogative-indefinite was favoured. Although it remains uncertain (while still acceptable, as discussed above) why that would be the case for the Germanic languages, in typological accordance with the grammaticalisation paths undergone by the Italic relative pronoun, it is plausible to assume that the element being reanalysed into an argumental complementiser in the Germanic languages was indeed the relative, and not the demonstrative⁷.

While, at this point, the general outline of the above-described grammaticalisation paths seems to be – at least on principle – fairly acceptable, the aim of this study is that of trying to shed some light on the multiple controversies pertaining to this topic.

3. Methodology

The linguistic corpus which has been compiled specifically for this study consists of about eight hundred sentences from the earliest attested stages of the main Germanic language. These sentences were mostly gathered from three different sources: the *Early Indo-European Online* repository by the University of Texas at Austin, the *RuneS* website, and the handbook of Germanic Philology *Old English and its Closest Relatives* by Orrin W. Robinson.

The prime criterion employed to select the relevant sentences was the presence within each one of them of at least one element which, according to the Present-Day English translation provided, could be rendered as *that*. At a later stage, the collected entries were further selected, by only keeping those which covered one of the functions that have been instantiated so far (i.e. demonstrative pronoun, demonstrative determiner, definite article, relative pronoun, argumental complementiser, and consecutive conjunction). Any token which, despite its function, did not have a relevant origin (e.g. Old

⁷ This hypothesis does, however, still need testing.

English *swaswa* is a consecutive conjunction meaning *so that*, but it clearly does not derive from any form of the Proto-Indo-European **tód*) was disregarded.

For what concerns the selection of the languages to analyse, as already discussed [cf. Section 1], the cardinal principle was to provide a balanced account of each Germanic branch, in order to make sure any statistic result would be significant. Beside Gothic (East Germanic branch) and Old Norse (North Germanic branch), Old English was thus chosen among the West Germanic languages as it is the most represented one. Furthermore, Old English happens to pose major issues especially as far as the grammaticalisation of relatives is concerned, but has not been investigated as much as other languages, like Old High German, have. Nonetheless, some runic inscriptions and a handful of illustrative sentences from other West Germanic languages (i.e. Old High German, Old Low Franconian, Old Frisian, Old Saxon, and Middle English) are also included in the corpus.

From a chronological point of view, the oldest dated entries in the corpus are taken from the few surviving runic inscriptions in Older Futhork, which date from the second half of the second century CE to the end of the seventh century. The corpus also features some interesting Younger Futhork (700-1325) and Futhork (700-1100) inscriptions. While it is debated whether Older Futhork should be defined as Proto-Norse or as a form of Common-Germanic (Robinson 1992), it is clear that Younger Futhork inscriptions are purely Scandinavian, whereas Futhork is considered as a form of Old English. Notably, the oldest runic inscriptions typically consist of one or two descriptive words and serve the purposes of “naming objects, naming the owners of objects, naming the carver of the runes, naming the person memorialised, or some combination of the above” (Robinson 1992: 95-96).

As for non-runic texts, the corpus covers quite an extensive period of time (roughly from the fourth century to the late fourteenth century) and a wide range of text genres. In particular, as far as Gothic is concerned, the passages which have been inserted in the corpus are taken from the Gospels in Wulfila’s Bible (fourth century) and from the *Skeirins* – a fifth century commentary on the Gospel of Mark. Religious content may however also be

found in some West Germanic entries, and especially in those taken from the following texts:

- An Old English poetic paraphrase of the *Genesis* written by an English monk around 1000;
- The Old Saxon alliterative epic poem *Heliand* (first half of the ninth century), where an account of the life of Jesus is provided, through the combination of all four Gospels;
- The ninth century Old High German eschatological poem *Muspilli*;
- The word-for-word interlinear translations from Latin of the Gospels and the Book of Psalms in, respectively, Old High German and Old Low Franconian⁸, for which no certain date can be provided;
- The Middle English historical and religious poem *Cursor Mundi* (1300).

With respect to religious texts in Old Norse, on the other hand, we are mainly referring to the myths of the Germanic heathendom which may be found in some excerpts of a 1270s manuscript of Eddic poetry which is included in the corpus, as well as in some passages from the Prologue of Snorri Sturluson's (1179-1241) prosaic *Edda*.

Besides religious content, however, a significant number of the entries in the corpus are taken from historical accounts, such as the Old English *Anglo-Saxon Chronicle* (years 755, 897, and 991) and their most recent realisation, i.e. the *Peterborough Chronicle* (year 1087) – of which some passages written in the Middle English period (year 1137) are also included. Similar accounts also existed in Old Norse literature and some examples may be found in sentences taken from an account of the Battle of Stamford Bridge in 1066 and in the *Íslendingabók*, or the *Book of Icelanders*, which was written by the priest Ari Þorgilsson (1067-1148) and dealt with the early history of Iceland.

Among Old Norse literary texts, we may also find the so-called “family sagas,” which usually narrate the story of the families that first settled in

⁸ While some scholars claim that this language may be called “Old Dutch,” Robinson (1992: 205) argues that it rather is “the oldest stage of the Limburgic dialect of Dutch, which differs substantially from the standard language.” For this reason, it is thus preferable to define this early phase as “Old Low Franconian.”

Iceland. This genre is represented in the corpus through passages from the *Brennu-Njáls Saga* (composed around 1250-127 by an unknown author) and from a fourteenth century manuscript of the *Egils Saga*, whose author might again be Snorri Sturluson. In addition to family sagas, however, the corpus also contains some excerpts from “legendary sagas,” like the *Grettis Saga* and the *Hrólfs saga* (both taken from fourteenth century manuscripts), which tell stories of heroes and warriors and whose genre greatly resembles that of epic poems such as the Old English *Beowulf*, of which some passages taken from a 1000 lone manuscript are provided.

The corpus also includes entries from two Old English elegiac poems from the tenth century *Exeter Book* (i.e. *The Seafarer* and *The Wanderer*), a lecture on the poet Caemdon by the Venerable Bede (671-735), and King Alfred’s (c. 848-899) account of *The Voyages of Ohthere and Wulfstan*, as well as some sentences taken from more recent sources, like the Old Frisian 1300 *Asegabook* (i.e. the legal code of the Frisians inhabiting the Rustringian region of Frisia) and Geoffrey Chaucer’s *Tales of Canterbury*, from the Middle English period (1387).

Once all the relevant sentences had been collected, the tokens that were translatable as *that* were analysed by part of speech and by the syntactic function they covered within the context of the sentence. Based on the syntactic function of the entries, the corpus was thus divided into four separate tables⁹: demonstratives, articles, relatives and conjunctions. All of the entries featured in these tables were ordered chronologically, based on the assumed dating of the manuscript from which they were taken, and a number of variables were included so as to provide a more detailed analysis of the tokens in question.

As it is evident from the selection included in the Appendix section, each entry is associated with a progressive number, in order to make any reference throughout the paper more accessible. The first few columns indicate the language and text from which the sentence was taken, as well as its dating. The last two columns, on the other hand, feature the translation of the sentence and the source from which the entry itself was collected. Although,

⁹ A selection of each of the four resulting tables is presented in the Appendix section of this paper.

more often than not, the translation was already provided in the consulted repositories, adjustments were made when a more literal translation was needed to render the structure of the original sentence more apparent. Finally, the primary sources that had been consulted to compile the corpus [cf. Primary sources in Bibliography] were associated with a letter and the relevant letter for each of the entries was thus inserted in the “Source” column.

However, there are some differences among the tables which are worth analysing separately. The “Demonstrative” table indicates the form from which the token came (e.g. *sa* comes from *sa*, *so*, *þata*), its semantic agreement by case, gender, and number and, finally, its syntactic function – i.e. demonstrative pronoun, demonstrative determiner, or person pronoun. In this table, the “Part of Speech” column is not included, as it was the same (i.e. demonstrative pronoun) for all of the entries.

Similarly, since the syntactic function was the same for all the entries in both the “Article” and “Relative” tables, only the origin of the token and the semantic agreement are indicated.

In conclusion, the “Conjunction” table includes both the syntactic function of argumental complementiser and that of consecutive conjunctions, and the entries are therefore analysed by origin of the token and syntactic function. While they were all uninflected forms and the semantic agreement was thus irrelevant, the verb by which the argumental clause is introduced was highlighted, in the case of argumental complementisers, as well as the possible presence of a consecutive structure (e.g. *so*+ADV/ADJ *that*), in the case of a consecutive conjunction.

4. Corpus analysis

4.1. Demonstratives

As already discussed, the Proto-Indo-European anaphoric pronoun **só-*, **sá-*, **tód* is the common ancestor of at least some of the demonstrative pronouns and determiners which we find in the Present-Day Germanic languages. As shown in Table 2, this is true for the earliest stages of the Germanic languages, as well. Although, in fact, the Gothic form **jáins** and the Old Norse *hann* and *hinn* do not derive from **só-*, **sá-*, **tód*, most of the demonstratives represented in the corpus are indeed recognised as its reflexes (e.g. **in bold**,

the Gothic *sa, so, þata*, the Old English *sē, sēo, ðæt* and the Old Norse *sá* and *sjá*).

Forms	Gothic			Old English		Old Norse				
	<i>sa, so, þata</i>	<i>jáins</i>	Tot.	<i>sē, sēo, ðæt</i>	Tot.	<i>hann</i>	<i>hinn</i>	<i>sá</i>	<i>sjá</i>	Tot.
No. Instances	35	3	38	28	28	55	2	135	1	198
Demonstrative Pronoun	13	0	13	12	12	8	2	82	1	93
Demonstrative Determiner	8	2	10	12	12	1	0	11	0	17
Third-Person Pronoun	14	1	15	4	4	46	0	42	0	88

Table 2. Distribution of deictics within the corpus.¹⁰

As it is evident from the table, in all of the Germanic branches, the demonstratives tended to function not only as deictic pronouns, which could then be used both cataphorically (7) and anaphorically (8), but also as demonstrative determiners (9) and third-person pronouns (10). While in most Present-Day Germanic languages this latter use has been dropped, the determiner function has been preserved and even consolidated.

- (7) GOT *Sa ist sunus meins sa liuba [...]*.
 this is son my the beloved¹¹
 ‘This is my son, the beloved.’

[cf. Appendix, no. 3]

- (8) GOT *Jah þatei gasahv jag gahausida þata weitwodeiþ.*
 and what saw.3SG and heard that testify.3SG
 ‘And what he saw and heard, that he testifies.’

[cf. Appendix, no. 4]

¹⁰ In line with the choices that have been made regarding the corpus, Table 2, as well as most of the tables included in this paper, only features instances from one language per Germanic branch (i.e. Gothic, Old English and Old Norse). Although, as already mentioned in Section 3, the intent behind this decision was to provide an even representation of each branch, this has not always been the case, since the linguistic corpora of the different languages are obviously not equivalent.

¹¹ Relevant elements are often not glossed for morpheme, as the relative information can be found either in the textual explanation of the examples or in the provided tables.

- (9) ON *Fyrir þá sök [...].*
for that reason
'For that reason.'

[cf. Appendix, no. 8]

- (10) EN *Mid þæm þæt manige [...] forðfēdon.*
among them (the) many died
'Among them, many [...] died'

[cf. Appendix, no. 6]

The aim of this section is therefore to provide an overview of how the demonstratives are represented within the corpus, while analysing each of the three Germanic branches separately and making some considerations on the evolution of the demonstratives systems in the different languages.

4.1.1. East Germanic

Table 3 accounts for the use of demonstratives in Gothic, in that each lemma instantiated in the demonstratives part of the corpus is analysed by case, number and gender and the number of occurrences per syntactic function is registered. The table also features information on the texts where such function was featured and the dating of those texts.

	Inflection			Demonstrative Pronoun		Demonstrative Determiner		Third-Person Pronoun		Tot.
				No.	Text, dating	No.	Text, dating	No.	Text, dating	
<i>ijos</i>	NOM	-	PL	0	-	0	-	1	<i>Gospels</i> , 4 th cent.	1
<i>jainaim</i> ¹²	DAT	M	PL	0	-	‡	<i>Gospels</i> , 4 th cent.	‡	<i>Gospels</i> , 4 th cent.	‡
<i>jainans</i> ¹³	ACC	M	PL	0	-	‡	<i>Gospels</i> , 4 th cent.	0	-	‡

¹² This form does not derive from the Proto-Indo-European *só-*, **sá-*, **tód*. Its attestations are therefore not relevant for the purposes of this study and they are not included in the sum total reported in the last row of Table 3.

¹³ This form does not derive from the Proto-Indo-European *só-*, **sá-*, **tód*. Its attestations are therefore not relevant for the purposes of this study and it is not included in the sum total reported in the last row of Table 3.

<i>sa</i>	NOM	M	SG	2	<i>Gospels</i> , 4 th cent.	1	<i>Gospels</i> , 4 th cent.	0	-	3
<i>soh</i>	NOM	F	SG	0	-	1	<i>Gospels</i> , 4 th cent.	1	<i>Gospels</i> , 4 th cent.	2
<i>þai</i>	NOM	M	PL	1	<i>Gospels</i> , 4 th cent.	1	<i>Gospels</i> , 4 th cent.	1	<i>Gospels</i> , 4 th cent.	3
<i>þaim</i>	DAT	M	PL	1	<i>Gospels</i> , 4 th cent.	0	-	5	<i>Gospels</i> , 4 th cent.	7
		N		1	<i>Gospels</i> , 4 th cent.	0	-	0	-	
<i>þamma</i>	DAT	M	SG	0	-	1	<i>Gospels</i> , 4 th cent.	1	<i>Gospels</i> , 4 th cent.	2
<i>þata</i>	NOM	N	SG	1	<i>Gospels</i> , 4 th cent.	0	-	1	<i>Gospels</i> , 4 th cent.	11
	ACC	N	SG	3	<i>Gospels</i> , 4 th cent.; <i>Skeireins</i> , 5 th cent.	2	<i>Gospels</i> , 4 th cent.	3	<i>Gospels</i> , 4 th cent.	
1				<i>Gospels</i> , 4 th cent.	0	-	0	-		
<i>þat(uh)</i>										
<i>þis</i>	GEN	N	SG	1	<i>Skeireins</i> , 5 th cent.	0	-	0	-	1
<i>þize</i>	GEN	N	PL	1	<i>Gospels</i> , 4 th cent.	0	-	0	-	1
<i>þo</i>	NOM	N	SG	1	<i>Skeireins</i> , 5 th cent.	0	-	0	-	4
		N	PL	0	-	0	-	1	<i>Gospels</i> , 4 th cent.	
	ACC	F	SG	0	-	1	<i>Skeireins</i> , 5 th cent.	0	-	
		N	PL	0	-	1	<i>Gospels</i> , 4 th cent.	0	-	
Tot.				13		8		14		35

Table 3. Distribution of Gothic deictics in the corpus.

The table provides a first confirmation of what has been discussed so far, in that – as expected – most of the instantiated lemmas are but inflected forms of the Gothic demonstrative *sa*, *sō*, **þata**, which originated from the Proto-In-

do-European *só-, *sá-, *tód and, just like it, expressed a *that*-deixis. Interestingly, the only exceptions are the forms *jainans* and *jainaim* which represent the distal-deixis inflected forms of *jáins*, *jaina*, *jainata* and both derive from the Proto-Germanic **jainaz*, **jenaz* – just as the Present-Day German distal demonstrative *jener*, *jene*, *jenes* does (Keretchashvili 2021). The very fact that this form exists in German and is in a complementary distribution with the proximal demonstrative *der*, *die*, *das* (i.e. a reflex of *só-, *sá-, *tód) seems to indicate that the demonstratives deriving from *jáins*, *jaina*, *jainata* and those deriving from *sa*, *sō*, **pata** might have been in a complementary distribution in Gothic, as well.

If this were the case, we could assume that the Gothic demonstrative forms deriving from *só-, *sá-, *tód were already fully grammaticalised as demonstrative determiners as early as in the fourth century. In fact, as results from Table 3, the inflected forms of *sa*, *sō*, **pata** function as demonstrative pronouns – used both cataphorically (7) and anaphorically (8) – as well as as demonstrative determiners, when in adjectival position. Moreover, the Gothic *sa*, *sō*, **pata** also appears to have largely been used as a third-person pronoun.

There does not appear to be a pattern by which certain forms alone might take on specific functions, while others could not. While, within the pool of sentences analysed in the corpus, some forms do not feature all of the functions taken into consideration, it is reasonable to assume that this is just imputable to the fact that the entries gathered for the corpus were too few and/or – having been randomly chosen – not representative enough.

Despite its being a closed-corpus language, Gothic still represents the optimal point of departure for this research as it is the earliest attested Germanic language. On the one hand, we can, in fact, assume a closer resemblance with the non-attested Proto-Germanic language and expect Gothic to thus provide us with invaluable insight on those earlier stages of the Germanic languages. On the other hand, however, we need to bear in mind that all the Gothic texts we have today are translations from Koine Greek and that, on these grounds, the written form of the Gothic language has been largely influenced and even shaped by it.

Nevertheless, seeing how Koine Greek demonstratives worked further confirms the patterns shown in the corpus. In Greek, in fact, the grammat-

icalisation of demonstratives into determiners had occurred already in the Classical period and was preserved throughout the Koine Greek time so that the demonstratives could be used indifferently as pronouns or adjectives.

While a third-person pronoun existed in Greek, however, it was rarely used and often replaced by the oblique cases of the demonstrative αὐτός, αὐτή, αὐτό (Rydberg-Cox 2021). This is not unlike what seems to have happened in Gothic, where – despite the existence of the “proper” third-person pronoun *is, eis, ita* (Keretchashvili 2021) – the demonstrative *sa, sō, þata* is often preferred.

4.1.2. West Germanic

With regards to the Old English demonstratives, as seen earlier on in this chapter, all of the instances of demonstratives in the corpus are inflected forms of *sē, sēo, ðæt* – i.e. an Old English reflex of the Proto-Indo-European **só-*, **sá-*, **tód*. In Old English *sē, sēo, ðæt* used to express a that-deixis, while today, after the dropping of the distal pronoun *yon* (from the Proto-Germanic **jainaz, jenaz*), its descendent *that, those* has taken over the distal-deixis.

					Demonstrative Pronoun	Demonstrative Determiner	Third-Person Pronoun	Tot.		
	Inflection				No	Texts, Dating	No		Texts, Dating	No
<i>ðæt</i>	ACC	N	SG	0	-	1	<i>The Voyages of Ohthere and Wulfstan</i> (848-899)	0	-	1
<i>þā</i>	ACC	-	PL	0	-	2	The Venerable Bede (671-735)	0	-	2
<i>þa</i>	ACC	F	SG	1	<i>Genesis</i> , 1000	1	<i>Genesis</i> , 1000	0	-	2
<i>þām</i>	DAT	M	SG	0	-	0	-	1	<i>Exeter Book</i> , 10 th cent.	1
<i>þam</i>	DAT	-	PL	0	-	2	Peterborough Chronicle, 1087	0	-	2
<i>þāra</i>	GEN	-	PL	1	<i>Anglo-Saxon Chronicle</i> (y. 897)	0	-	1	The Venerable Bede (671-735)	2

<i>þǣm</i>	DAT	-	PL	0	-	1	<i>Anglo-Saxon Chronicle (y. 897)</i>	0	-	4
			N	SG	0	-	1	<i>The Voyages of Ohthere and Wulfstan (848-899)</i>	0	
	NOM	N	SG	1	<i>Anglo-Saxon Chronicle (y. 897)</i>	1	<i>Anglo-Saxon Chronicle (y. 897)</i>	0	-	
<i>þæs</i>	GEN	N	SG	4	<i>Anglo-Saxon Chronicle (y. 755, 897); Beowulf's Prologue, 1000</i>	0	-	0	-	4
<i>þæt</i> ¹⁴				3	<i>The Voyages of Ohthere and Wulfstan (848-899); Anglo-Saxon Chronicle (y. 991); Beowulf's Prologue, 1000</i>	0	-	1	<i>The Voyages of Ohthere and Wulfstan (848-899)</i>	10
				3	<i>The Venerable Bede (671-735); Anglo-Saxon Chronicle (y. 755); Exeter Book, 10th cent.</i>	3	<i>The Voyages of Ohthere and Wulfstan (848-899); Genesis, 1000</i>	0		
Tot.								3		28

Table 4. Distribution of Old English deictics in the corpus.

As shown in Table 4, the few instances of demonstratives featured in the corpus were almost equally used both pronominally and adjectivally – thus showing that also in Old English the grammaticalisation of demonstrative pronouns into determiners had already taken place.

Conversely, there are only three sentences in which the demonstrative is used as a third-person pronoun and that leaves the question open about

¹⁴ Spelling variation of *þæt*.

whether the pool of entries was again too small and not representative enough, or whether those were merely sporadic uses which had never really become established. Either way, based on some extracts from the *Canterbury Tales*, this specific use seems to have already been dropped by the Middle English period.

As for the other West Germanic languages, as reported in the corpus through few but significant examples, the Old Saxon *that*, the Old Frisian *thi*, *thiu*, *thet* and the Old High German *daz*, *thaz* – as well as their inflected forms – all functioned as both demonstrative pronouns and determiners with deictic function when used in adjectival position. In addition, Old High German featured a second demonstrative pronoun, *thiz*, which had an emphatic function.

Old Low Franconian, on the other hand, stands out for being the only West Germanic language in which the grammaticalisation of determiners appears to have not yet occurred. The demonstrative *thie*, therefore, was probably only used as an anaphoric pronoun with deictic function (11).

- (11) OLF *Ginathi in uuarheide sina uue sal thia suocan?*
 grace and truth his who shall those question
 ‘His grace and truth, who shall question those?’

[cf. Appendix, no. 9]

Nevertheless, this seems merely due to a delay in the rise of the category of definiteness in the earlier stages of the Dutch language, as in Present-Day Dutch it is perfectly acceptable to use demonstratives in adjectival position (12).

- (12) NL *Die zwarte fiets is de mijne.*
 that black bike is the mine
 ‘That black bike is mine.’

Among the Germanic languages analysed so far, however, Present-Day Dutch is also the only one in which a demonstrative still acts as the singular neuter third-person pronoun. This is, however, merely due to the assimilation of the two forms which occurred in Modern times.

4.1.3. North Germanic

As shown in Table 2, the Old Norse demonstratives system is composed of four main pronouns, with distinct origins and/or meanings. The most represented pro-

noun in the corpus is *sá, sú, þat*, which is a reflex of the Proto-Indo-European **só-*, **sá-*, **tód* and expresses a *that*-deixis. A similar form, of the same derivation, is constituted by the proximal demonstrative pronoun *sjá/þessi, sjá/þessi, þetta*, which, however, only appears in one instance throughout the corpus. Another common demonstrative in the corpus is *hinn, hin, hið*, which – despite its derivation from the Proto-Germanic distal demonstrative pronoun **jainaz*, like the Gothic *jáins, jaina, jainata* (Keretchashvili 2021) – is assumed to have lacked any specific spatial deixis and was often used as the second referent in constructions like “the one ... the other,” where it was contrasted with another referent marked by either *sá* or *sjá* (Krause and Slocum 2005). Finally, another way to express the demonstrative function in Old Norse was through the third-person pronoun *hann, hon, þat*. This pronoun was really a demonstrative as its declension constituted a suppletive system, where its masculine and feminine singular forms were built from the demonstrative *hann* (again originating from the Proto-Germanic **jainaz*), whereas the neuter singular and the plural of all genders stemmed from the *þ*-forms found in the declension of the demonstrative *sá, sú, þat* (Krause and Slocum 2005).

Given this premise, it is therefore clear why so many entries featuring *hann* would have the function of third-person pronoun. Nonetheless, because of its direct origin from **só-*, **sá-*, **tód*, the main focus in this chapter will remain the demonstrative *sá, sú, þat* (it is, in fact, impossible to draw any conclusion based on the single instance of the other reflex of **só-*, **sá-*, **tód*, i.e. *sjá/þessi, sjá/þessi, þetta*).

	Inflection				Demonstrative Pronoun	Demonstrative Determiner	Third-Person Pronoun	Tot.		
	No	Texts, Dating		No	Texts, Dating		No		Texts, Dating	
<i>hinn</i> ¹⁵	NOM	M	SG	z	<i>Poetic Edda</i> , 1270s; <i>Hrólfs saga</i> , late 14 th cent.	0	-	0	-	z

¹⁵ This form does not derive from the Proto-Indo-European *só-*, **sá-*, **tód*. Its attestations are therefore not relevant for the purposes of this study and they are not included in the sum total reported in the last row of Table 5.

sá	NOM	M	SG	2	Poetic Edda, 1270s	2	Battle of Stamford Bridge (y. 1066); Hrólfs saga, late 14 th cent.	3	Poetic Edda, 1270s	7
sú	NOM	F	SG	2	Edda (1179-1241); Brennu-Njáls Saga (1250-1275)	0	-	0	-	2
þá	ACC	F	SG	0	-	1	Edda (1179-1241)	0	-	4
		M	PL	2	Battle of Stamford Bridge (y. 1066); Poetic Edda, 1270s	0	-	1	Brennu-Njáls Saga (1250-1275)	
þann	ACC	M	SG	2	Poetic Edda, 1270s; Hrólfs saga, late 14 th cent.	4	Íslendingabók (1067-1148); Edda (1179-1241); Poetic Edda, 1270s;	2	Poetic Edda, 1270s; Egils Saga, late 14 th cent.	8
þat	NOM	N	SG	6	Íslendingabók (1067-1148); Brennu-Njáls Saga (1250-1275); Poetic Edda, 1270s; Grettis Saga, 14 th cent.	0	-	6	Edda (1179-1241); Brennu-Njáls Saga (1250-1275); Poetic Edda, 1270s; Hrólfs saga, late 14 th cent.	50
	ACC	N	SG	32	Íslendingabók (1067-1148); Edda (1179-1241); Brennu-Njáls Saga (1250-1275); Poetic Edda, 1270s; Hrólfs saga, late 14 th cent.; Egils Saga, late 14 th cent.	2	Poetic Edda, 1270s;	4	Edda (1179-1241); Hrólfs saga, late 14 th cent.	

<i>þau</i> ¹⁶	NOM	N	PL	2	Íslendingabók (1067-1148); Brennu-Njáls Saga (1250-1275)	0	-	1	Poetic Edda, 1270s	4	
	ACC	N	PL	1	Poetic Edda, 1270s	0	-	0	-		
<i>þeim</i>	DAT	M	PL	0	-	0	-	1	Battle of Stamford Bridge (y. 1066)	1	
<i>þær</i>	NOM	F	PL	0	-	0	-	4	Poetic Edda, 1270s	5	
	ACC	F	PL	1	Edda (1179-1241)	0	-	0	-		
<i>þeim</i>	DAT	M	SG	1	Poetic Edda, 1270s	0	-	2	Poetic Edda, 1270s	21	
				1	Poetic Edda, 1270s	0	-	7	Íslendingabók (1067-1148); Edda (1179-1241); Brennu-Njáls Saga (1250-1275); Hrólfs saga, late 14 th cent.		
		N	PL	F	1	Poetic Edda, 1270s	0	-	0		-
				1	Edda (1179-1241)	0	-	8	Edda (1179-1241); Brennu-Njáls Saga (1250-1275); Poetic Edda, 1270s; Hrólfs saga, late 14 th cent.		

¹⁶ This form does not derive from the Proto-Indo-European *só-*, **sá-*, **tód*. Its attestations are therefore not relevant for the purposes of this study and they are not included in the sum total reported in the last row of Table 5.

<i>þeir</i> ¹⁷	NOM	M	PL	6	Íslendingabók (1067-1148); <i>Edda</i> (1179-1241); <i>Brennu-Njáls Saga</i> (1250-1275); <i>Poetic Edda</i> , 1270s; <i>Hrólfs saga</i> , late 14 th cent.	±	Íslendingabók (1067-1148);	40	<i>Battle of Stamford Bridge</i> (y. 1066); Íslendingabók (1067-1148); <i>Edda</i> (1179-1241); <i>Brennu-Njáls Saga</i> (1250-1275); <i>Poetic Edda</i> , 1270s; <i>Grettis Saga</i> , 14 th cent.; <i>Hrólfs saga</i> , late 14 th cent.	47
<i>þeir</i> ¹⁸	GEN	M	PL	0	-	0	-	2	<i>Edda</i> (1179-1241); <i>Hrólfs saga</i> , late 14 th cent.	6
		N	PL	0	-	0	-	4	<i>Edda</i> (1179-1241); <i>Brennu-Njáls Saga</i> (1250-1275)	
<i>þeiri</i>	DAT	F	SG	0	-	0	-	1	<i>Brennu-Njáls Saga</i> (1250-1275)	1
<i>þeirri</i>	DAT	F	SG	1	Íslendingabók (1067-1148)	0	-	0	-	1

¹⁷ This form does not derive from the Proto-Indo-European *só-*, **sá-*, **tód*. Its attestations are therefore not relevant for the purposes of this study and they are not included in the sum total reported in the last row of Table 5.

¹⁸ This form does not derive from the Proto-Indo-European *só-*, **sá-*, **tód*. Its attestations are therefore not relevant for the purposes of this study and they are not included in the sum total reported in the last row of Table 5.

<i>þess</i>	GEN	M		0	-	1	Poetic Edda, 1270s	0	-	4
		N	SG	3	Íslendingabók (1067-1148); Egils Saga, late 14 th cent.; Hrólfs saga, late 14 th cent.	0	-	0	-	
<i>þet</i>	NOM	N	SG	3	Battle of Stamford Bridge (y. 1066)	0	-	1	Battle of Stamford Bridge (y. 1066)	6
	ACC	N	SG	2	Battle of Stamford Bridge (y. 1066)	0	-	0	-	
<i>því</i>	DAT	N	SG	23	Battle of Stamford Bridge (y. 1066); Íslendingabók (1067-1148); Edda (1179-1241); Poetic Edda, 1270s; Grettis Saga, 14 th cent.; Egils Saga, late 14 th cent.; Hrólfs saga, late 14 th cent.	1	Poetic Edda, 1270s	1	Hrólfs saga, late 14 th cent.	25
Tot.				83		11		41		135

Table 5. Distribution of Old Norse deictics in the corpus.

As reported in Table 5, the most represented function in the corpus, as far as the inflected forms of *sá*, *sú*, **þat** are concerned, is that of demonstrative pronoun – and unsurprisingly so, since that was the original function even in Proto-Indo-European. What, on the other hand, seems noteworthy is that, in comparison, its adjectival use as a demonstrative determiner is significantly under-represented. Although, as already mentioned, this could just be due to the random selection of entries, the difference in the number of occurrences is large enough to at least hypothesise that this function had not yet been fully established at the time in which

the texts included in the corpus were written. This latter hypothesis seems to be supported also by the fact that, in the Present-Day North Germanic languages, the demonstrative is equally used both as a pronoun and as an adjective.

Moreover, as already seen in Old English and especially in Gothic, even the “proper” demonstrative *sá, sú, þat* was often used as a third-person pronoun – a use which has been dropped in Present-Day Scandinavian languages, where, conversely, the descendants of the Old Norse pronoun *hann* have stabilised as the only third-person pronouns.

4.2. Articles

As discussed before, in Proto-Indo-European, the relationship between the demonstrative and the name to which it referred is believed to have been an appositional one (Quiles and López-Menchero 2011). While we cannot reconstruct the grammatical category of the article in Proto-Indo-European, as well as in the earliest attested stages of ancient Indo-European languages such as Homeric Greek, we do however find definite articles, for instance, in later stages of Greek itself. We may therefore still assume that – probably once the structure (anaphoric demonstrative pronoun + noun) and its meaning had stabilised – the nature of the demonstrative changed from appositional to adjectival, thus giving rise to the category of determiners.

The [+definiteness] feature, which generally distinguishes definite articles, could be expressed by the appositional demonstrative already in Proto-Indo-European (at least in contexts where it was salient) and has been the triggering element which gave rise to the category of determiners. It was thus only when, at a later stage, the determiner lost its [-deictic] feature – while preserving its [+definiteness] one – that it became a definite article. For this reason, articles may be considered as less marked than demonstrative determiners, as they lack the [-deictic] feature. Nonetheless, it is still unknown at what point in the history of the individual Germanic languages this renovation came into being.

Skrzypek, Piotrowska, and Jaworski (n.d.: 26-27) maintain that, whereas “there does seem to exist a difference between the demonstrative, in which the definite article originates, and the definite article itself,” such difference is “not clear-cut, and therefore the stages of grammaticalisation are not discrete.” They argue, in fact, that the development of the demonstrative

“forms a continuum from a purely deictic element to a distinct definite article” (Skrzypek, Piotrowska, and Jaworski n.d.: 27). From this perspective, we may thus expect some features belonging to the demonstrative as well as some of those proper to the article to have, at some point, coexisted within the grammaticalising demonstrative.

The main purpose of this chapter is therefore that of comparing the corpus data which have been annotated as either demonstrative determiners or definite articles [Table 6] and trying to decide, for each one of the languages taken into account, whether or not this latter function was already present and what its degree of grammaticalisation was at the time instantiated. Moreover, the origin of the North Germanic definite articles will be discussed, as well as the definite articles systems in the Present-Day Germanic languages.

Forms	Gothic	Old English	Old Norse					Tot.
	<i>sa, so, þata</i>	<i>sē, sēo, ðæt</i>	<i>hann</i>	<i>hinn</i>	<i>sá</i>	<i>sjá</i>	Tot.	
Demonstrative Determiners [cf. Section 2]	8	12	1	0	11	0	17	37
Definite Articles	52	12	1	6	12	1	22	86

Table 6. Distribution of determiners in the corpus.

4.2.1. East Germanic

Table 7 features the instances of Gothic demonstrative determiners and definite articles found in the corpus and analyses each lemma based on its inflection and occurrence.

	Inflection			Definite Article		Demonstrative Determiner		Tot
				No	Texts, Dating	No	Texts, Dating	
	<i>jainaim</i> ¹⁹	DAT	M	PL	0	-	‡	Gospels, 4 th cent.

¹⁹ This form does not derive from the Proto-Indo-European *só-*, **sá-*, **tód*. Its attestations is therefore not relevant for the purposes of this study and it is not included in the sum total reported in the last row of Table 7.

<i>jainans</i> ²⁰	ACC	M	PL	0	-	±	<i>Gospels</i> , 4 th cent.	±
<i>sa</i>	NOM	M	SG	6	<i>Gospels</i> , 4 th cent.	1	<i>Gospels</i> , 4 th cent.	7
<i>soh</i>	NOM	F	SG 6 <i>Gospels</i> , 4 th cent.	0	-	1	<i>Gospels</i> , 4 th cent.	7
<i>so</i>				0	-			
<i>pai</i>	NOM	M	PL	6	<i>Gospels</i> , 4 th cent.	1	<i>Gospels</i> , 4 th cent.	7
<i>paim</i>	DAT	M	PL	3	<i>Gospels</i> , 4 th cent.	0	-	3
<i>pamma</i>	DAT	M	SG	10	<i>Gospels</i> , 4 th cent.	1	<i>Gospels</i> , 4 th cent.	11
<i>pana</i>	ACC	M	SG	4	<i>Gospels</i> , 4 th cent.	0	-	4
<i>pans</i>	ACC	M	PL	3	<i>Gospels</i> , 4 th cent.	0	-	3
<i>pata</i>	NOM	M	SG	1	<i>Gospels</i> , 4 th cent.	0	-	4
	ACC	M	SG	1	<i>Gospels</i> , 4 th cent.	2	<i>Gospels</i> , 4 th cent.	
<i>pe</i>	INSTR	N	SG	1	<i>Skeireins</i> , 5 th cent	0	-	1
<i>pis</i>	GEN	M	SG	2	<i>Gospels</i> , 4 th cent.	0	-	3
		N		1	<i>Gospels</i> , 4 th cent.	0	-	
<i>pizai</i>	DAT	F	SG	1	<i>Gospels</i> , 4 th cent.	0	-	1
<i>pize</i>	GEN	M	PL	0	-	0	-	2
		N	PL	2	<i>Gospels</i> , 4 th cent.	0	-	
<i>pizos</i>	GEN	F	SG	1	<i>Gospels</i> , 4 th cent.	0	-	1

²⁰ This form does not derive from the Proto-Indo-European *só-*, **sá-*, **tód*. Its attestations is therefore not relevant for the purposes of this study and it is not included in the sum total reported in the last row of Table 7.

þo	ACC	F	SG	2	Gospels, 4 th cent.	1	Skeireins, 5 th cent.	5
		N	PL	1	Skeireins, 5 th cent.	1	Gospels, 4 th cent.	
þos	ACC	F	PL	1	Gospels, 4 th cent.	0	-	1
Tot.				52		8		60

Table 7. Distribution of Gothic determiners in the corpus.

As a first indication, all of the instances of definite articles are inflected forms of the Gothic demonstrative *sa*, *sō*, **þata**, while, in two occurrences, the demonstrative determiner is expressed through the already discussed forms *jainans* and *jainaim* [cf. Section 2]. Granted that the amount of data is not sufficient to draw any relevant conclusion on this point, it is safe to say that the reflex of **so-*, **sa-*, **tód* is at least the prevalent option in the formation of articles. This strong connection between the demonstrative and the article is, moreover, confirmed by the case of Greek, where the demonstrative **ὁδε**, **ἦδε**, **τόδε** is composed of the definite article **ὁ**, **ἡ**, **τό** (< PIE **so-*, **sa-*, **tód*) and the enclitic suffix **-δε**. Greek, in fact, was also the first Indo-European language to grammaticalise definiteness and to thus introduce definite articles. Indeed, we can find examples of overlapping distribution between the original demonstrative **ὁ**, **ἡ**, **τό** and the homographic definite article already in the *Odyssey*, as well as in the latest books of the *Iliad* (Chantraine 1958:165).

Despite the large difference in the number of occurrences of Gothic demonstrative determiners and definite articles in the corpus, in the absence of any reason to suppose otherwise, we may assume that any inflected form of the Gothic *sa*, *sō*, **þata** could be used to express both functions. Furthermore, written Gothic was a translational language based on a source language (i.e. testamentary Greek) which did feature definite articles and it is thus reasonable to presume that the target language would use them, as well.

Since, however, the identical surface of demonstrative determiners and definite articles is a confusing factor, in order to identify any difference of use between the two, we need to have a closer look at the meanings expressed within the specific contexts in which they occur.

Example (13) presents two occurrences of demonstratives used in adjectival position, both showing a clear [+deictic] feature: in the first case, in fact, “*jainans*” refers to some specific days, while, in the second case, the deictic function is even more explicit, in that “*soh gilstrameleins*” (“this taxing”) constitutes an anaphoric reference to the “*gameljan*” (“being taxed”) introduced in the previous sentence.

- (13) GOT *Wa þan in dagans jainans, urrann gagrefts fram kaisara*
 happened and in days those rose decree from Caesar
Agustau, gameljan allana midjungard. Soh þan gilstrameleins frumista
 Augustus, be.taxed all world this and taxing first
warþ at [...].
 happened when
 ‘And it happened in those days (that) a decree from Caesar Augustus
 (stating that) all the world should be taxed rose. And this taxing first
 happened when [...]’

[cf. Appendix, no. 1-2]

- (14) GOT *was-uh þan nehva pasxa, so dulþs Iudaie.*
 was-and then near Passover the feast Jew.GEN.PL
 ‘And then the Passover, the feast of the Jews, was near.’

[cf. Appendix, no. 11]

As for example (14), however, it is beyond doubt that the determiner *so* is to be interpreted as an article: we could, in fact, translate it as either “a feast of the Jews” or “the feast of the Jews,” depending on the importance of the feast from a cultural point of view, but we could never render it as “that feast of the Jews” unless we make the reference more specific, e.g. “that feast of the Jews during which they commemorate their liberation from slavery in Egypt” (in this latter case, the [+deictic] feature could only be motivated by emphasis, as mentioned in the previous example). However, in the absence of such addition and of any emphatic purpose in the passage taken from the corpus, it is clear that the function of the determiner *so* in this sentence can only be that of an article.

Even these few examples, in fact, seem to provide sufficient evidence to conclude that the signature features of definite articles were indeed present in the Gothic determiner *sa*, *sō*, ***þata*** and that, when it was used as such, it lacked, on the other hand, the [-deictic] feature which is proper to the

demonstrative. It is thus reasonable to suppose that, not unlike the “poly-functional *that*” in Present-Day English, the Gothic surface-identical *sa*, *sō*, **þata** could function as a demonstrative pronoun, demonstrative determiner, third-person pronoun, and as a definite article. Although some degree of ambiguity is still present at times, we could argue that, for the most part, each function had specific features and contexts of use.

4.2.2. West Germanic

According to corpus data relative to Old English, both the entries annotated as demonstratives and those annotated as definite articles are all inflected forms of *sē*, *sēo*, *ðæt* and Table 8 provides a compared analysis of their occurrences in the corpus, with specific attention to the lemmas involved.

Unfortunately, however, the instances gathered in the corpus, as far as definite articles and demonstrative determiners are concerned, is too limited to make any relevant consideration and the fact that Table 8 shows that the definite article is only expressed through three forms (i.e., in **bold**, *ðā*, *þāra*, and *þæt*) is just due to their under-representation within the corpus. All the inflected forms of the Old English *sē*, *sēo*, *ðæt* served, in fact, both as the demonstrative pronoun/determiner *that*, *those*, and as the definite article *the*. Although *sē*, *sēo*, *ðæt* did have a peculiar declension, which may be considered as suppletive, since the various forms are quite difficult to predict, the only true difference between the demonstrative forms and the definite article ones, is that the article was unstressed and the masculine nominative form was thus spelled as *se* – rather than *sē*.

	Inflection				Definite Article		Demonstrative Determiner		Tot.
					No.	Texts, Dating	No.	Texts, Dating	
<i>ðæt</i>	ACC	N	SG	0	-	1	<i>The Voyages of Ohthere and Wulfstan</i> (848-899)	1	
<i>ðā</i>	NOM	-	PL	3	<i>The Voyages of Ohthere and Wulfstan</i> (848-899); <i>Beowulf</i> , 1000	0	-	3	

<i>þā</i>	ACC	-	PL	0	-	2	The Venerable Bede (671-735)	2
<i>þā</i> ²¹	ACC	F	SG	0	-	1	<i>Genesis</i> , 1000	1
<i>þam</i>	DAT	-	PL	0	-	2	Peterborough Chronicle, 1087	2
<i>þāra</i>	GEN	-	PL	1	<i>Anglo-Saxon Chronicle</i> (y. 897)	0	-	1
<i>þāem</i>	NOM	N	SG	0	-	1	<i>Anglo-Saxon Chronicle</i> (y. 897)	3
	DAT	N	SG	0	-	1	<i>The Voyages of Ohthere and Wulfstan</i> (848-899)	
		-	PL	0	-	1	<i>Anglo-Saxon Chronicle</i> (y. 897)	
<i>þæt</i> ²²	NOM	N	SG	3	<i>The Voyages of Ohthere and Wulfstan</i> (848-899); <i>Anglo-Saxon Chronicle</i> (y. 897)	0	-	10
	ACC	N	SG	4	The Venerable Bede (671-735); <i>The Voyages of Ohthere and Wulfstan</i> (848-899); <i>Exeter Book</i> , 10 th cent.	3	<i>The Voyages of Ohthere and Wulfstan</i> (848-899); <i>Genesis</i> , 1000	
Tot.				11		12		23

Table 8. Distribution of Old English determiners in the corpus.

The examples which can be found in the corpus seem to show that there indeed existed some difference – albeit subtle – in the contexts in which the

²¹ Spelling variation of *þā*.

²² Spelling variation of *ðæt*.

Old English *sē*, *sēo*, *ðæt* could be used, based on whether it expressed deixis (15) or definiteness (16).

- (15) OE [...] *hīe wēron micle swiþor gebrocede on þām þrim gēarum*
 they were much more afflicted in those three years
mid cēapes cwilde ond monna.
 by cattle.GEN.SG pestilence and man.GEN.PL
 ‘They were much more afflicted in those three years by pestilence of
 cattle and men.’

[cf. Appendix, no. 5]

- (16) OE [...] *hē forlēt þæt hūs þæs gebēorscipes* [...]
 he left the house the.GEN.SG feast
 ‘He left the house of the feast’

[cf. Appendix, no. 12]

There are, however, also some more ambiguous instances, where a “partial distributional overlap” may be identified, as in the example reported by Crisma (2020: 39–40), taken from the *Catholic Homilies* (c. 990) by the English abbot Ælfric, where the phrase “*ðæt beclýsede geat*” could be equally rendered as either “*the locked gate*” or “*that locked gate*.”

- (17) OE *Eft Ezechiel se witega geseah on his witegunge an belocen geat*
 then Ezechiel the prophet saw in his prophecy a locked gate
on Godes huse. & him cwæð to sum engel. Þis geat
 in God.GEN house and him said to some angel this gate
ne bið nanum men geopenod ac se hlaford ana færd
 NEG is no.DAT men.DAT opened but the Lord one goes
in þurh þæt geat. & eft ut færd & hit bið
 in through that gate and back out goes and it is
belocen on ecnyse. Ðæt beclýsede geat
 locked in eternity The/That locked gate
on Godes huse getacnode þone halgan mægðhad þære
 in God.GEN house betokened the holy virginity the.GEN
eadigan Marian
 blessed.GEN Mary.GEN
 ‘Also Ezechiel the prophet saw in his prophecy a locked gate in God’s
 house, and an angel said to him, “This gate shall be opened to no man, for
 the Lord alone will go in through that gate and then back out, and it shall

be locked for ever”. The/That locked gate in the house of God signified the holy virginity of the blessed Mary.’

(Crisma 2020:39-40)

This ambiguity between the surface-identical demonstrative determiners and definite articles has often led scholars to wonder if there even existed a proper definite article (and, by extension, the category of definiteness altogether) in Old English. As Allen (2019) claims, in fact, we must not attempt to impose Present-Day categories on any earlier stage of the languages.

Before we can question the status of the Old English definite article, however, we need to understand why the case of English would be more prominent than that of the other languages where such ambiguity can be found – like, for instance, Gothic and Present-Day German.

As already discussed, the Gothic *sa*, *sō*, *þata* could be considered, at least to some extent, as an ambiguous determiner, in that it might not always be possible to distinguish when it served as a deictic marker from when it was used as a definiteness one. Unlike Old English, however, as we have already mentioned, Gothic is known to have had direct and massive contact with a language which already featured grammaticalised definiteness (i.e. Greek) and it is thus more straightforward to assume that it, too, presented definite articles.

Moreover, since the Gothic language died out quite soon, one can only speculate about whether at some point the two functions might have split into distinct forms. On the other hand, a compelling reason why the status of the Old English article is object of debate is indeed the presence, in Middle English, of a new form of the article (i.e. the invariant *þe*²³), which was “phonologically and semantically distinct from the distal demonstrative²⁴,” which by that time had, in turn, developed into the invariant *ðat* and specialised as the distal demonstrative (Crisma 2020: 38). It is, in fact, the very existence of two distinct forms in later stages of the English lan-

²³ As we will see in [Section 4.3.2], in Old English, *þe* was also a relative particle.

²⁴ Definite articles which are distinct from the demonstrative are defined as “discrete definite articles” (Crisma 2020: 44).

guage that have brought the invariant *þe* to be regarded as the “proper” definite article.

- (18) OE *Ðe wrecce men of þe land*
 the wretched men of the land
 ‘The wretched men of the land’

[cf. Appendix, no. 14-15]

Although its role as a definiteness marker is beyond doubt (18), however, it does not constitute sufficient reason to state that the previous form (i.e. *sē, sēo, ðæt*) did not cover that function, as well. The Present-Day German *der, die, das* is, in fact, both a demonstrative pronoun and a definite article²⁵ and neither of its functions is called into question. Therefore, the issue at hand is not really the overlapping distribution of the two functions, but rather whether grammaticalised definiteness had already been accomplished in the Old English period (Crisma 2020).

In her research, Crisma (2020) confirms that evidence of grammaticalised definiteness could be found in Old English prose (eighth century), but not in the Old English poetry and concludes that this grammaticalisation process must have occurred during the Old English period, centuries before the split of the two forms. On these grounds, although we cannot pinpoint the exact moment in which definiteness was grammaticalised, we may generally say that, roughly after the eighth century, when *sē, sēo, ðæt* was used to indicate definiteness, from a syntactical point of view, it was indeed a definite article.

While the morphological split could be “a case of parallel decay in form and meaning” (Crisma 2000: 76), in that both forms, *þæt* and *þe*, lost their inflectional properties and, respectively, their [-definiteness] and [-deictic] features, from a typological point of view, the grammaticalisation of definiteness is a process which has occurred in a significant number of languages, including non-Indo-European ones.

As for the other West Germanic languages, the Old Saxon *that*, the Old Frisian *thi, thiu, that*, and the Old High German *daz, thaz* were all examples of “near-articles,” in that they were demonstrative pronouns which, just like the Gothic *sa, sō, þata* and the Old English *sē, sēo, ðæt*, could be used as definite

²⁵ “A demonstrative pronoun with article function is called a “near-article” (Crisma 2020: 44).

articles. Besides *daz*, *thaz*, however, Old High German also featured a second determiner which could also be used as a relative pronoun, i.e. *der*, *ther*.

Finally, as already mentioned, Old Low Franconian did not yet feature grammaticalised definiteness and therefore did not have definite articles. On the other hand, Present-Day Dutch does present a definite article (i.e. *de*, *het*²⁶), also because, as we have seen, it is not a direct ancestor of Dutch.

4.2.3. North Germanic

Table 9 presents a comparison of the Old Norse entries from the corpus which have been annotated as either demonstrative determiners or as definite articles. While, in the corpus, the demonstrative determiner is almost always expressed by the Old Norse demonstrative *sá*, *sú*, *þat* (except for one single instance where the third-person pronoun *hann*, *hon*, *þat* is used), the representation of the definite article is more heterogeneous. We can find, in fact, a restricted number of cases in which it is expressed by the demonstratives *hinn*, *hin*, *hið* and *sjá/þessi*, *sjá/þessi*, *þetta*.

Nevertheless, as shown in Table 9, the inflected forms of *sá*, *sú*, *þat* do prevail also in the realisation of the definite article, while the other options analysed occur far too sporadically for us to make any significant consideration in those regards.

				Definite Article		Demonstrative Determiner		
Inflection				No.	Texts, Dating	No.	Texts, Dating	Tot.
<i>Hinn</i> ²⁷	NOM	F	SG	±	<i>Edda</i> (1179-1241)	0	-	2
	ACC	M	SG	±	<i>Battle of Stamford Bridge</i> (y. 1066)	0	-	

²⁶ In Present-Day Dutch, while, from a functional point of view, the article *het* is a reflex of **tód*, such derivation is incorrect from an etymologic perspective. In a previous stage to Present-Day Dutch, in fact, there existed a neuter article deriving from **tód*, i.e. *dat*. However, such article underwent aphesis (thus being reduced to *t*) and was, over time, assimilated to another pronoun which had undergone aphesis, as well, i.e. the third-person pronoun *het*.

²⁷ This form does not derive from the Proto-Indo-European *só-*, **sá-*, **tód*. Its attestations are therefore not relevant for the purposes of this study and they are not included in the sum total reported in the last row of Table 9.

<i>hinn</i> ²⁸	NOM	M	SG	3	<i>Battle of Stamford Bridge</i> (y. 1066); <i>Hrólfs saga</i> , late 14 th cent.	0	-	3
<i>hinum</i> ²⁹	DAT	M	PL	±	<i>Hrólfs saga</i> , late 14 th cent.	0	-	±
<i>sá</i>	NOM	M	SG	1	<i>Edda</i> (1179-1241)	2	<i>Battle of Stamford Bridge</i> (y. 1066); <i>Hrólfs saga</i> , late 14 th cent.	3
<i>þá</i>	ACC M	F	SG	0	-	1	<i>Edda</i> (1179-1241)	2
		PL	1	<i>Edda</i> (1179-1241)	0	-		
<i>þann</i>	ACC	M	SG	0	-	4	<i>Íslendingabók</i> (1067-1148); <i>Edda</i> (1179-1241); <i>Poetic Edda</i> , 1270s;	4
<i>þat</i>	NOM	N	SG	1	<i>Egils Saga</i> , late 14 th cent.	0	-	3
	ACC	N	SG	0	-	2	<i>Poetic Edda</i> , 1270s	
<i>þæim</i>	DAT	M	SG	1	<i>Battle of Stamford Bridge</i> (y. 1066)	0	-	1
<i>þeim</i>	DAT	M	SG	1	<i>Poetic Edda</i> , 1270s;	0	-	1
<i>þeir</i> ³⁰	NOM	M	PL	0	-	±	<i>Íslendingabók</i> (1067-1148);	±

²⁸ This form does not derive from the Proto-Indo-European *só-*, **sá-*, **tód*. Its attestations are therefore not relevant for the purposes of this study AND THEY are not included in the sum total reported in the last row of Table 9.

²⁹ This form does not derive from the Proto-Indo-European *só-*, **sá-*, **tód*. Its attestations are therefore not relevant for the purposes of this study AND IT IS not included in the sum total reported in the last row of Table 9.

³⁰ This form does not derive from the Proto-Indo-European *só-*, **sá-*, **tód*. Its attestations are therefore not relevant for the purposes of this study and it is not included in the sum total reported in the last row of Table 9.

<i>þeira</i> ³¹	GEN	N	PL	‡	Íslendingabók (1067-1148)	0	-	‡
<i>þeiri</i>	DAT	F	SG	1	<i>Edda</i> (1179-1241)	0	-	1
<i>þenn</i>	ACC	M	SG	2	<i>Battle of Stamford Bridge</i> (y. 1066)	0	-	3
<i>þenna</i>	ACC	M	SG	1	<i>Battle of Stamford Bridge</i> (y. 1066)	0	-	
<i>þess</i>	GEN	M	SG	1	Íslendingabók (1067-1148)	1	<i>Poetic Edda</i> , 1270s	4
		N		2	<i>Hrólfs saga</i> , late 14 th cent.	0	-	
<i>því</i>	DAT	N	SG	1	Íslendingabók (1067-1148)	1	<i>Poetic Edda</i> , 1270s	2
Tot.				13		11		24

Table 9. Distribution of Old Norse determiners in the corpus.

From a closer look at the sentences exemplified in the corpus, we can also see that, in addition to the significant instability in the formation of the definite article, there are virtually no instances in which the [+definiteness] feature indisputably prevails over the [+deictic] feature.

The slight difference in meaning between the forms annotated as demonstrative determiners (19) and those annotated as definite articles (20) is, in fact, not relevant enough to suppose that the demonstrative *sá, sú, þat* could indeed serve as a definiteness marker with no [-deictic] feature.

- (19) ON [...] Óláfr inn Digri gærði skýrt at hverr maðr skyldi gjalda [...]

Olaf the thick made definite that each man should pay

nema konur

except women

³¹ These forms do not derive from the Proto-Indo-European *só-, *sá-, *tód and are therefore not relevant for the purposes of this study. These instances are not included in the sum total reported in the last row of Table 9.

eða þeir menn er hann næmi frá.
 or those men whom he exempts from
 ‘Olaf the Thick made definite that each man should pay [...] except women or those men whom he exempts.’

[cf. Appendix, no. 7]

(20) ON *Kenndo þér þenn hinn myckla meðr þæim blá kyrtli oc hin faghra hialm [...]?
 know you that this.one noble with the blue kirtle and the impressive helmet*

‘Do you know that noble man with the blue kirtle and impressive helmet [...]?’

[cf. Appendix, no. 13]

On the other hand, we cannot rule out the possibility that the category of definiteness was grammaticalised, by virtue of the fact that Old Norse did feature a postpositive definite article (21), i.e. *inn, in, it* – whose declension followed that of the demonstrative *hinn, hin, hið* (Krause and Slocum 2005).

(21) ON [...] *Svá kom at þeir týndu guðs nafni ok víðast um veröld-ina fansk eigi sá maðr er deili kunnir á skapara sínum.*
 so occurred that they forgot God.GEN name and widely around world-ART could.be.found not DET man who details knew of creator his

‘So it happened that they forgot God’s name and in most places around the world a man could not be found who knew the details of his creator.’

(*Snorra Edda*, Krause and Slocum 2005)

While *hinn, hin, hið* expressed distal deixis and derived from an Indo-European pronoun combining variants of the stems **ko-* and **eno-* (cf. the Greek distal demonstrative ἐκεῖνος), the article *inn, in, it* was a Proto-Scandinavian innovation (Stroh-Wollin 2020). This article could be found in some *Older Fuþark* runic inscriptions [cf. *hino* in Appendix, no. 10], according to Stroh-Wollin (2020), derived from an anciently inherited demonstrative based on the stem *en-* (< Proto-Indo-European **eno-*), which had originally been strengthened through the addition of a Proto-Germanic root expressing proximal deixis (**hī-* or **hi-* < Proto-Indo-European **ko-*). As we can thus notice, this account represents typological evidence of the grammaticalisation of demonstratives into definite articles, regardless of the stem from which the demonstrative came.

As for the Present-Day Scandinavian languages, as already seen [cf. Section 2], they all present the suffixal form derivating from *-inn*, *-in*, *-it* as their sole or – at least – primary definiteness marker. While, in fact, Danish and Swedish do also feature a free-standing article whose form is compatible with a possible origination from *sá*, *sú*, *þat*, Norwegian and Icelandic only use the suffix. However, since the suffix *-inn*, *-in*, *-it* is assumed to have originated from the above-mentioned article *inn*, *in*, *it* (Stroh-Wollin 2020), we could argue that all the North Germanic languages have undergone the grammaticalisation path by which a demonstrative became a definite article (Skrzypek, Piotrowska, and Jaworski n.d.).

4.3. Relatives

Based on the relatives system in the Present-Day Germanic languages, we may assume that one of the West Germanic relatives (i.e. the *d-Pronomen*) derived from the Proto-Indo-European **so-*, **sa-*, **tód*, while there is no trace of such relative form in the Scandinavian languages. What we do not know, however, is how and why this development came into being, as well as when it occurred and what languages it involved.

Through an in-depth analysis of some salient entries from the corpus, this section thus tries to verify whether the demonstrative had already been grammaticalised into a relative in the early stages of the Germanic languages and, if so, in which languages. The main criterion that will be used to test the relevant sentences is the relationship between the main clause and the supposed relative one. In order to consider a relative as such, in fact, the two clauses must form at least a correlative structure (i.e. either a *diptyque normal* or a *diptyque inverse*), in that, in any other case, the two clauses would just be coordinates and the element heading the second clause would have no reason to be interpreted as anything different from a mere anaphoric demonstrative.

FORMS	GOTHIC		OLD ENGLISH		OLD NORSE		
	<i>sa, so, þata</i>	<i>sa + ei</i>	<i>sē, sēo, ðæt</i>	<i>þe</i>	<i>sa + er</i>	<i>er</i>	<i>sem</i>
Inflected	6	17	5	0	8	0	0
Non-inflected	0	0	0	31	0	46	7
Tot.	24		37		61		

Table 10. Distribution of relative elements in the corpus.

At a first look, all the Gothic, Old English, and Old Norse tokens featured in the corpus which have been annotated as relatives belong to three categories: demonstratives, relative particles, or a combination of both [Table 10]. All the demonstratives instantiated are reflexes of *so-, *sa-, *tód, while the so-called “combinations” are compound forms composed of a demonstrative element (again from *so-, *sa-, *tód) and a relative particle. As for their specific forms, as could be expected, all the demonstratives and the compounds are inflected, while the relative particles are clearly indeclinable.

4.3.1. East Germanic

As shown in Table 10, based on the corpus data, Gothic seems to have expressed relatives either through the simple demonstrative *sa*, *so*, *þata* or through the compound form made of *sa*, *so*, *þata* itself and the relative particle *ei*. In addition, Table 11 features an analysis of the specific lemmas involved and provides information on their occurrence within the corpus.

	Inflection			Demonstrative		Demonstrative + Relative Particle			Tot.
				No	Texts, Dating	No	Texts, Dating	Form	
<i>sa</i>	NOM	M	SG	1	<i>Gospels</i> , 4 th cent.	5	<i>Gospels</i> , 4 th cent.	<i>saei</i>	6
<i>so</i>	NOM	F	SG	0	-	3	<i>Gospels</i> , 4 th cent.	<i>sei</i>	3
<i>þai</i>	NOM	M	PL	1	<i>Gospels</i> , 4 th cent.	1	<i>Skeireins</i> , 5 th cent.	<i>þaiei</i>	2
<i>þaim</i>	DAT	M	PL	1	<i>Gospels</i> , 4 th cent.	0	-	-	1
<i>þamma</i>	DAT	M	SG	2	<i>Gospels</i> , 4 th cent.	0	-	-	2
<i>þana</i>	ACC	M	SG	1	<i>Gospels</i> , 4 th cent.	0	-	-	1
<i>þata</i>	NOM	N	SG	0	-	4	<i>Gospel</i> , 4 th cent.	<i>þatei</i>	4
<i>þizai</i>	DAT	F	SG	0	-	1	<i>Gospel</i> , 4 th cent.	<i>þizaiei</i>	1

þo	ACC	F	SG	0	-	1	Gospels, 4 th cent.	þoei	2
		N	PL	0	-	1	Skeireins, 5 th cent.	þoei	
þos	ACC	F	SG	0	-	1	Gospels, 4 th cent.	þozei	1
Tot.				6		17			23

Table 11. Distribution of Gothic relative elements in the corpus.

According to Krause and Slocum (2006), Gothic had no independent pronoun which could function specifically as a relative and the relative marker function was thus covered by the particle *ei*³², when “conjoined to the demonstrative.” This compound is described as the relative pronoun for the third person, while, to create first and second-person pronouns the demonstrative was replaced by the appropriate form of the personal pronoun (e.g. *ik+ei*, *þu+ei*). In addition, Krause and Slocum (2006) argue that, just like in Present-Day English, the relative pronoun always “derives its case from its function in the relative clause.”

As previously stated, however, the goal of this corpus-based analysis of relatives is to verify whether the demonstrative pronoun derived from **so-*, **sa-*, **tōd* had already been grammaticalised into a relative in the early stages of the Germanic languages. To this aim, we must disregard the above-mentioned compounds, in that in such forms the relative function is expressed by the particle and there is no indication that the demonstrative could be anything different from an originally free-standing antecedent of the relative, which has later been conjoined to the relative particle.

Conversely, we will need to focus on the few cases in which the simple demonstrative is used and, for once, the low number of instances might indeed be an extremely significant indicator. Let us thus consider each of the six cases individually.

The first observation we can make on these passages is that in (22) and (23) the supposed relative clause features non-finite verb forms (i.e., more

³² Although its etymology is unclear, *ei* is presumed to have derived from the Proto-Indo-European anaphoric stem **i-* (Keretchashvili 2018).

specifically, two participles which could thus be considered as adjectives), whereas in (24), (25), (26), and (27) it does not present a verb at all. In either case, in the absence of a proper verb, the supposed relative cannot be defined as a clause. Nonetheless, the following sentences will be analysed in an attempt to further prove that, although the demonstrative in **bold** might be interpreted and translated in Present-Day English as a relative pronoun, it functioned as a determiner in Gothic.

- (22) a. GOT *usgeisnodedun þan allai þai*
 were.astonished.3PL and all.NOM.M.PL DEM.NOM.M.PL
hausjandans is ana frodein jah andawaurdjam is.
 hear-PRSP.NOM.M.PL him at wisdomand answers his
 ‘And all those hearing him were astonished at his wisdom and answers.’
 [cf. Appendix, no. 16]

- b. GR *ἐξίσταντο δὲ πάντες οἱ*
 were.astonished.3PL then all.NOM.M.PL DEM.NOM.M.PL
ἀκούοντες αὐτοῦ [...]
 hear-PRSP.NOM.M.PL him
 ‘Then all those hearing him were astonished.’

(Luke 2:47)

- (23) a. GOT *Iesu sokeiþ Nazoraiu þana ushramidan.*
 Jesus seek.2PL of.Nazareth DEM.ACC.M.SG crucify.PP.ACC.M.SG
 ‘You seek Jesus of Nazareth, the crucified.’
 [cf. Appendix, no. 17]

- b. GR *Ἰησοῦς ζητεῖτε τὸν Ναζαρηνοῦ τὸν*
 Jesus seek.2PL DEM.ACC.M.SG of.Nazareth DEM.ACC.M.SG
εσταυρωμένον
 exposed.PP.ACC.M.SG
 ‘You seek Jesus the Nazarene, the exposed’

(Mark 16:6)

As just mentioned, since “*hausjandans*” (22) and “*ushramidan*” (23) are verbal adjectives, the demonstrative pronoun which precedes and agrees with them could be described as a determiner³³ heading the appositional determiner

³³ As seen in Section 4.1.1 and Section 4.2.1, Gothic demonstratives were “near-articles” as they could indeed function both as demonstrative determiners and as definite articles.

phrases “*ḡai hausjandans*” and “*ḡana ushramidan.*” Interestingly, but perhaps also unsurprisingly, the Greek model (22b, 23b) from which these passages have been translated present the exact same constructions.

On the other hand, while with sentences (24), (25), (26), and (27) we could still be in the presence of appositional determiner phrases (there is, in fact, no evident correlative structure), we might assume that the verb *to be* is simply implied and that a relative clause is thus indeed present.

- (24) GOT [...] *atta* *izwar sa* *ufar himinam.*
 father.NOM.M.SG your DEM.NOM.M.SG above heaven.DAT.M.SG
 ‘[...] your father, the one that is above heavens.’
 [cf. Appendix, no. 18]

However, while this reasoning could hold for sentence (24), where the supposed relative pronoun – which, nonetheless agrees with its antecedent – could also be seen as deriving its case from the relative itself, it cannot be accepted with regards to sentences (25), (26) and (27). As we can see, in fact, the demonstrative always agrees with the noun it follows and does not derive its case from its function in the supposed relative clause.

- (25) GOT [...] *ib jainaim ḡaim uta in gajukom*
 but DEM.DAT.M.PL DEM.DAT.M.PL without in parables
allata wairḡib
 all become.3SG
 ‘[...] But to them – those without (it) – all is done in parables.’
 [cf. Appendix, no. 19]
- (26) GOT [...] *attin izwaramma ḡamma in himinam.*
 father.DAT.M.SG your DEM.DAT.M.SG in heaven.DAT.M.SG
 ‘[...] Your father, the one in heaven.’
 [cf. Appendix, no. 20]
- (27) GOT [...] *bidei du attin ḡeinamma ḡamma*
 pray.imp.2sg to father.DAT.M.SG thy DEM.DAT.M.SG
in fulḡsrja,
 in secret.DAT.N.SG
 ‘[...] Pray to your father, the one that is in secret.’
 [cf. Appendix, no. 21]

One final and conclusive proof of this can be provided by example (28) – the natural continuation of sentence (27) – in which a proper relative clause is present (there is, in fact, the verb “*saihvīþ*”) and, consequently, the proper relative pronoun *saei* is used.

- (28) GOT [...] *jah atta þeins saei saihvīþ in fulhsnja*
 and father thy REL.NOM.M.SG see.3SG in secret.DAT.N.SG
usgibþ þus in bairhtein
 reward.3SG thee in brightness
 ‘[...] And your father who sees in secret, rewards you in brightness.’
 [cf. Appendix, no. 22]

Based on this analysis, it could thus be argued with a fair degree of certainty that Gothic demonstratives were not grammaticalised into relatives and that the only relative marker was indeed the compound form featuring the relative particle *ei*.

4.3.2. West Germanic

As far as Old English is concerned, Table 12 shows that its relative marker appears to have been expressed either by the relative particle *þe* or by an inflected demonstrative pronoun. However, differently from Gothic and Old Norse, the Old English relative particle *þe* is believed to have derived, itself, from the instrumental case of the demonstrative *sē, sēo, ðæt* (Keretchashvili 2018). However, based on the inflection of second-person pronoun (Keretchashvili 2018), its origin from the dative/instrumental singular form seems to be acceptable, as well.

	Inflection			POS	Texts, Dating	Tot.
<i>ðe</i> ³⁴	UNINFLECTED			Relative Particle	<i>Anglo-Saxon Chronicle</i> (y. 991); <i>Genesis</i> , 1000	2
<i>sē</i>	NOM	M	SG	Demonstrative Pronoun	<i>Anglo-Saxon Chronicle</i> (y. 755); <i>Beowulf</i> , 1000	2

³⁴ Spelling variation of *þe*.

<i>þæt</i>	NOM	N	SG	Demonstrative Pronoun	<i>The Voyages of Othhere and Wulfstan</i> (848-899); <i>Anglo-Saxon Chronicle</i> (y. 991)	2
	ACC	N	SG	Demonstrative Pronoun	<i>Genesis</i> , 1000	1
<i>þe</i>	UNINFLECTED			Relative Particle	The Venerable Bede (671-735); <i>Anglo-Saxon Chronicle</i> (y. 755, 897); <i>The Voyages of Othhere and Wulfstan</i> (848-899); <i>Exeter Book</i> , 10 th cent.; <i>Peterborough Chronicle</i> , 1087	29
<i>þone</i>	ACC	M	SG	Demonstrative Pronoun	<i>The Voyages of Othhere and Wulfstan</i> (848-899)	1
Tot.						39

Table 12. Distribution of Old English relative elements in the corpus.

As for the demonstrative, it only headed a limited number of supposed relative sentences in Old English and the ones included in the corpus will be individually analysed.

In examples (29), (30), and (31), although a finite verb is present in each supposed relative clause, the relation of the two clauses is not of a correlative type. The clause headed by the demonstrative can, in fact, be interpreted as either an apposition (29) or an attribute (30, 31) of the noun phrase to which it refers and, as a result, unlike correlated and subordinate clauses, they could stand on their own.

- (29) OE [...] *alle būtan ānum, sē wæs þæs aldormonnes godsunu.*
 all but one, DEM.NOM.M.SG was his noble godson
 ‘[...] all but one – that one was his noble godson.’

[cf. Appendix, no. 23]

- (30) OE *þonne is ān port on sūðewardum þām lande,*
 there is a port in the.south.of DEM.DAT.N.SG land
þone man hætt Scīringes hēal.
 DEM.ACC.M.SG one calls Skiringssal
 ‘There is a port in the south of that land – that (land) one calls Skiringssal.’

[cf. Appendix, no. 24]

- (31) OE *Geworhton* *ðā* | *Wedra* *lēode*
 built the of.Weders people
hlāw on *hliðe*, | *sē* *wæs* *hēah* *ond* *brād*, [...]
 mound upon hill DEM.NOM.M.SG was high and wide
 ‘The people of the Weders built a mound upon the hill – that (mound)
 was high and wide.’

[cf. Appendix, no. 26]

As for example (32), it is composed of one single clause and “*eall þæt*” is thus probably to be analysed as a syntagma used to expand the indefinite pronoun *all*.

- (32) OE [...] and *eall þæt* to *fæsle* *frea* *ælmīhtig*
 and all DEM.ACC.N.SG for progeny lord almighty
habban *wolde*
 hav would
 ‘and everyone (=all those) the Lord almighty would have for progeny.’

[cf. Appendix, no. 27]

The same could be said of sentence (33), where we can also find a second demonstrative functioning as a resumptive element, which might indicate that, even though the model of a correlative construction seems to be present, the structure had not yet been grammaticalised.

- (33) OE *Eal þæt* *his man* *āþer* *oððe* *ettan* *oððe* *erian* *mæg*,
 all DEM.NOM.M.SG a man either or graze or plough can
þæt *lið* *wið* *ðā* *sæ*.
 DEM.NOM.M.SG extends alongside the sea
 ‘Everything (all that) a man can either graze or plough – that extends
 alongside the sea.’

[cf. Appendix, no. 25]

While it is thus clear that the inflected demonstrative was not a relative in Old English, the status of the relative *þe* is, to this day, a central issue in Germanic studies.

- (34) OE *he wæs milde þam* *goum* *mannum* *þe* *god lufedon*.
 he was gentle DEM.DAT.PL good man.DAT.PL REL God love.3PL
 ‘He was gentle with those good men who love God.’

[cf. Appendix, no. 28]

As example (34) – among many others which can be found in the corpus – seems to indicate, the particle *þe* probably was a relative marker, just like the Gothic (and, as we will see, the Old Norse) relative particles. However, what is not clear is whether it was indeed a demonstrative which had been grammaticalised into a relative or something else entirely. During the Middle English period, in fact, it took over the function of the definite article, and was replaced as a relative marker by the form *þat* (Lee 2006) – which, again, derived from the demonstrative *sē, sēo, ðæt*. Furthermore, the Middle English *þat*, as well as the Old English *þæt*, could also be used as argumental complementisers and this is one of the main reasons why we could argue that the West Germanic *d-Pronomen* did not directly originate from the demonstrative, but rather from the complementiser itself³⁵.

Following on this hypothesis and on Bacskai-Atkari's study (2019), it would appear reasonable to maintain that the English *d-Pronomen* *that* was neither a decategorised pronoun nor a relative particle, but rather a relative complementiser.

As for the other West Germanic languages, they all expressed what in the corpus has been annotated as a relative through forms deriving from the demonstrative. These forms are either uninflected ones as the Old Saxon *that* (which also functions as a complementiser) and the Old Frisian relative particle *ther*, or inflected, as the Old Low Franconian *thie*, and the Old High German *thie*, *thar* and *der*, *ther*. However, none of these forms functioned as a true relative marker.

The sole exception was, in fact, the Old Frisian *thi*, *thiu*, *thet* (35), but we must, however, consider that this passage is taken from a fourteenth century manuscript, which is thus more recent than most of the texts included in the corpus.

- (35) OF *ther hi on eskriuin hede tha tian bodo, tha*
 there he on written had the ten commandments DEM.ACC.PL
skolde hi lera tha Israheliska folke.
 should he teach the Israhelite folk
 'On there he had written the ten commandments – the ones he should
 teach the Israelite folk.'

[cf. Appendix, no. 29]

³⁵ This hypothesis will be discussed in detail in Section 5 of this paper.

4.3.3. North Germanic

Finally, in Old Norse, as we can see from Table 13, there are two main types of relative markers: the relative particles *er*³⁶ and *sem*³⁷ and the compound forms composed of the demonstrative *sá, sú, þat* and the relative particle *er* itself.

	Inflection			POS	Texts, Dating	Tot.
<i>er</i> ³⁶	UNINFLECTED			Relative Particle	<i>Battle of Stamford Bridge</i> (y. 1066); <i>Íslendingabók</i> (1067-1148); <i>Edda</i> (1179-1241); <i>Brennu-Njáls Saga</i> (1250-1275); <i>Poetic Edda</i> , 1270s; <i>Egils Saga</i> , late 14 th cent.; <i>Grettis Saga</i> , 14 th cent.; <i>Hrólfs saga</i> , late 14 th cent.	46
<i>sás</i>	ACC	M	SG	Demonstrative <i>sá</i> + Relative Particle <i>er</i>	<i>Poetic Edda</i> , 1270s	1
<i>sem</i> ³⁹	UNINFLECTED			Relative Particle	<i>Hrólfs saga</i> , late 14 th cent.	6
				(Demonstrative Antecedent +) Relative Particle	<i>Battle of Stamford Bridge</i> (y. 1066)	±
<i>þærs</i>	NOM	F	PL	Demonstrative <i>sá</i> + Relative Particle <i>er</i>	<i>Poetic Edda</i> , 1270s	1

³⁶ The Old Norse relative particle *er* derives from the Proto-Germanic masculine third-person pronoun **iz* (Keretchashvili 2018).

³⁷ The Old Norse relative particle *sem* has uncertain origins, possibly from the Proto-Germanic adjective **saman*, meaning *same* (Keretchashvili 2018). The main difference between *er* and *sem* is that the latter was used especially in the later texts.

³⁸ This form does not derive from the Proto-Indo-European *só-*, **sá-*, **tód*. Its attestations are therefore not relevant for the purposes of this study and they are not included in the sum total reported in the last row of Table 9.

³⁹ This form does not derive from the Proto-Indo-European *só-*, **sá-*, **tód*. Its attestations are therefore not relevant for the purposes of this study and they are not included in the sum total reported in the last row of Table 9.

<i>þanns</i>	ACC	M	SG	Demonstrative <i>sá</i> + Relative Particle <i>er</i>	<i>Poetic Edda</i> , 1270s	3
<i>þats</i>	ACC	N	SG	Demonstrative <i>sá</i> + Relative Particle <i>er</i>	<i>Poetic Edda</i> , 1270s	2
<i>þeims</i>	DAT	M	SG	Demonstrative <i>sá</i> + Relative Particle <i>er</i>	<i>Poetic Edda</i> , 1270s	1
Tot.						8

Table 13. Distribution of Old Norse relative elements in the corpus.

According to Krause and Slocum (2005), the relative particles were often preceded by a form of *sá*, *sú*, *þat* which could alternatively belong to the relative clause (when compounded with the particle *er*), or to the preceding clause – thus bearing no relation to the case represented by the particle. In either case, the demonstrative is always an antecedent of the relative and not a relative marker itself. Since there is, moreover, no instance of use of the simple demonstrative as a relative, we might conclude that, based on corpus data, Old Norse did not have a grammaticalised relative pronoun which derived from the demonstrative *sá*, *sú*, *þat*.

Although this account would seem to be confirmed also by the distribution of the relatives in the Present-Day Scandinavian languages – where, in fact, no relative *d-Pronomen* is present – we cannot completely rule out the possibility that the grammaticalisation of the demonstrative into a relative might indeed have happened at a later stage than Old Norse and that the resulting *d-Pronoun* was then dropped some time before the Present-Day stage of the languages.

A possible confirmation in this direction would be the already mentioned account by Vigfusson (1874:29), in which he claims that the Icelandic archaic argumental complementiser *at* (which probably derived from the demonstrative *sá*, *sú*, *þat*) could also be used as a less frequent alternative of the relative particle *er*.

4.4 Conjunctions

This section intends to analyse the conjunctions which are believed to have originated, at least in some Present-Day Germanic languages, from the Proto-Indo-European demonstrative **só-*, **sá-*, **tód* – and, more specifically, the argumental complementiser and the consecutive conjunction.

Given the identity of form among demonstratives, relatives and conjunctions [cf. Section 2], the derivation of the latter from the first seems to be rather uncontroversial as far as the West Germanic languages are concerned. On the other hand, however, the origin of the Old Norse conjunction *at* and its reflexes across the Present-Day Scandinavian languages is, as previously anticipated, still object of debate.

Furthermore, since the previous chapter showed that relatives did not seem to be grammaticalised in the early stages of any of the Germanic languages, it is particularly interesting to verify whether argumental complementisers and consecutive conjunctions were.

This chapter will thus analyse the forms annotated in the corpus as either argumental complementisers or consecutive conjunctions and try to establish their origin (especially as far as the North Germanic languages are concerned) and their degree of grammaticalisation.

As shown in Table 14, we can see at a first look that all of the forms instantiated (except for the Gothic compound relative *sa + ei*) are invariant ones. While we recognise the relative particles *ei* and *sem* in, respectively, Gothic and Old Norse, and the now uninflected Old English demonstrative *ðæt/þæt*, we can also find some other forms which are worth discussing individually in the dedicated sections.

FORMS	GOTHIC				OLD ENGLISH	OLD NORSE		Tot.
	<i>ei</i>	<i>þatei</i>	<i>þei</i>	<i>sa + ei</i>	<i>ðæt/þæt</i>	<i>at</i>	<i>sem</i>	
ARG	3	9	0	2	17	48	1	80
CONS	6	0	3	0	6	28	0	43
Tot.	23				23	77		121

Table 14. Distribution of conjunctions in the corpus.

Despite the individual differences among the different languages, however, all argumental complementisers do share, as a common feature, the context in which they can be found. Such context is given by the verbs introducing them, which all belong to specific semantic categories. Although the general label of *verba dicendi* and *sentiendi* – which has been used so far – is still applicable, as it is evident from Table 15, there are in fact many more categories involved.

Semantic Category	Verbs	GOT			OE	ON		Tot.
		<i>ei</i>	<i>þatei</i>	<i>sa + ei</i>	<i>ðæt/ þæt</i>	<i>at</i>	<i>sem</i>	
Predicate Adjectives	be	0	0	0	0	1	0	1
	be best	0	0	0	0	1	0	1
	be far	0	0	0	0	1	0	1
	be fitting	0	0	0	1	0	0	1
	be likely	0	0	0	0	2	0	2
Verba Dicendi	claim	0	0	0	0	1	0	1
	say	0	5	0	8	5	0	18
	tell	0	1	0	0	2	0	3
Verba Desiderandi	desire	0	0	0	0	1	0	1
	want	0	0	0	0	1	0	1
Verba Existimandi	decide	0	0	0	0	1	0	1
	determine	0	0	0	0	1	0	1
	know	0	1	0	2	3	0	6
	mean	0	0	0	0	1	0	1
	reason	0	0	0	0	1	0	1
	settle (on)	0	0	0	0	1	0	1
	suspect	0	0	0	0	1	0	1
understand	0	0	0	0	2	0	2	
Verba Iubendi	command	1	0	0	0	0	0	1

Verba Sentiendi	hear	0	1	0	2	0	0	3
	see	0	0	2	0	8	0	10
	seem	0	0	0	1	0	1	2
	watch	0	0	0	0	1	0	1
Others	come about	0	0	0	0	1	0	1
	discover	0	0	0	0	2	0	2
	happen	0	0	0	0	2	0	2
	make clear	0	0	0	0	1	0	1
	offer	0	0	0	2	0	0	2
	pray	1	0	0	0	0	0	1
	strike	0	0	0	1	0	0	1
	reveal	0	0	0	1	0	0	1
write	1	1	0	0	0	0	2	
Non-ARG		0	0	0	0	6	0	6
Tot.		3	9	2	17	48	1	80

Table 15. Semantic categories of verbs introducing argumental complementisers in the corpus.

In conclusion, while there are still many issues to be addressed with reference to argumental complementisers, consecutive conjunctions seem to be relatively unproblematic, in that their grammaticalisation is considered to be an extension of the meaning of the complementiser, when in the contexts of a consecutive construction, such as *so+ADJ/ADV...that*.

4.4.1. East Germanic

As reported in Table 16, Gothic argumental and consecutive clauses were headed by the relative particle *ei* – which was occasionally compounded with other forms, mainly taken from the demonstrative *sa, so, þata*.

However, while in a restricted number of instances (in the corpus there are only two) this compound was indeed inflected and corresponded to the relative pronoun *sa+ei* [cf. Section 4.3.1], the compound *þatei* is an invariant form (composed of the nominative neuter singular form of the demonstra-

the *þata* and the particle *ei*) which seems to have been a subordinating conjunction in its own right (Krause and Slocum 2006). Similarly, the form *þei* was itself an invariant compound functioning as a subordinating conjunction, but no part of it seems to have derived from the demonstrative: in fact, its *þ*- stem (to which the particle *ei* was added) derived from the Proto-Germanic pronominal stem **þa*- (Keretchashvili 2018).

	Inflection			POS	Argumental Complementiser		Consecutive Conjunction		Tot.
					No.	Texts, Dating	No.	Texts, Dating	
<i>ei</i>	UNINFLECTED			Particle	3	<i>Gospels</i> , 4 th cent.	6	<i>Gospels</i> , 4 th cent.; <i>Skeireins</i> 5 th cent.	9
<i>þammei</i>	DAT	N	SG	Relative Pronoun (> <i>saei</i>)	2	<i>Gospels</i> , 4 th cent.	0	-	2
<i>þatei</i>	UNINFLECTED			Conjunction	9	<i>Gospels</i> , 4 th cent.	0	-	9
<i>þei</i>	UNINFLECTED			Conjunction	0	-	3	<i>Gospels</i> , 4 th cent.	3
Tot.					14		9		23

Table 16. Distribution of Gothic conjunctions in the corpus.

Based on this account, we can thus deduce that in Gothic the simple demonstrative was never used as a conjunction, which clearly indicates that it never did grammaticalise into either of the forms we are considering in this chapter (i.e. argumental complementiser and consecutive conjunction).

Nonetheless, we may notice a certain degree of similarity between relative and argumental clauses, as they are both headed by conjunctions or pronouns which contain the relative marker *ei*.

4.4.2. West Germanic

As for the Old English argumental and consecutive clauses, they seem to have all been headed by an invariant form of the demonstrative *sē*, *sēo*, *ðæt* – and,

more specifically, the nominative neuter singular form *ðæt*, whose alternative orthography *þæt* is also attested, and often preferred [Table 17].

	Inflection	POS	Argumental Complementiser		Consecutive Conjunction		Tot.
			No.	Texts, Dating	No.	Texts, Dating	
<i>ðæt</i>	UNINFLECTED	CONJ	1	<i>The Voyages of Ohthere and Wulfstan</i> (848-899)	0	-	23
<i>þæt</i>			16	<i>Anglo-Saxon Chronicle</i> (y. 755, 897); <i>The Voyages of Ohthere and Wulfstan</i> (848-899); <i>Exeter Book</i> , 10 th cent.; <i>Beowulf</i> , 1000	6	<i>The Venerable Bede</i> (671-735); <i>The Voyages of Ohthere and Wulfstan</i> (848-899); <i>Exeter Book</i> , 10 th cent.; <i>Beowulf</i> , 1000	
Tot.			17		6		

Table 17. Distribution of Old English conjunctions in the corpus.

Since all the forms instantiated in the corpus are thus derived from the Proto-Indo-European **só-*, **sá-*, **tód*, we will analyse some examples in order to verify the degree of grammaticalisation into a conjunction of the demonstrative itself.

- (36) OE *For þon cnyssað nū*
 therefore strikes now
heortan geþōhtas, | þæ ic hēan strēamas,
 heart.GEN.F.SG thought.ACC.M.PL ARG I humble.ADJ stream.ACC.M.PL
sealtýþa gelāc | sylf cunnige;
 seawaves.GEN.F.PL tumult myself test.1SG
 ‘Therefore it strikes now the thoughts of the heart that I test myself the humble streams, the tumult of sea waves.’

[cf. Appendix, no. 31]

As it is evident from example (36), the Old English *þæt* does seem to be fully grammaticalised as an argumental complementiser, and – as mentioned at the beginning of this chapter – even when the matrix verb is not strictly a *verbum dicendi* or *sentiendi*, just like in the Present-Day variety of English.

However, while the grammaticalisation path of the argumental complementiser into a consecutive conjunction might be fairly straightforward

on principle, it does not seem to have occurred at this stage of Old English. Throughout the corpus, in fact, the conjunction is almost never found in the typical consecutive context (i.e. *so+ADJ/ADV...that*) and seems to cover more of an explicative function (37) – thus resulting as something which is still in between an argumental complementiser and a proper consecutive conjunction (as a matter of fact, in the use of the demonstrative we can still identify a trace of the *diptyque inverse* correlative structure mentioned in Section 2). On the other hand, in the one instance in which it is accompanied by the construction *so+ADJ*, it does indeed function as a fully fledged consecutive conjunction (38).

- (37) OE *þā hē þæt þā sumre tīde dyde,*
 and he DEM.ACC.N.SG then one time did.3SG
þæt hē forlēt þæt hūs þæs gebēorscipes [...]
 CONJ he left ART.ACC.N.SG house ART.GEN.M.SG feast.GEN.M.SG
 ‘And then one time he did this – that he left the house of the feast.’
 [cf. Appendix, no. 30]

- (38) OE *For þon nis þæs mōdwlanc mon ofer eorþan, [...]*
 because is.not so proud.ADJ man on earth
þæt hē ā his sǣfōre | sorga næbbe,
 CONJ he never his sea.travel anxiety has
 ‘because not a man on earth is so proud that he never has his sea travel anxiety.’
 [cf. Appendix, no. 31]

As we have already seen in the previous chapter [cf. Section 4.3.2], the Old English *þæt* went through a vocalic sound change ($\text{æ} > \text{a}$) and, over the Middle English period transformed first into *þat* and then into the Present-Day form *that*. On the other hand, the fact that the Middle English relative complementiser took the form of *þat*, as well, while the particle *þe* was dropped (it lost the relative marker function, but transformed into the definite article) seem to imply that the relative did grammaticalise afterwards and on the basis of the grammaticalisation of the argumental complementiser.

Furthermore, the fact that the relative *d*-Pronoun did not derive directly from the demonstrative seems to be confirmed by the forms used in Old High

German, as well. The Old High German demonstratives and complementisers were in fact surface-identical (i.e. *daz*, *thaz*), while the relative had a slightly different form (i.e. *der*, *ther*). In addition, as we can see, Old High German did preserve both a common gender (for masculine and feminine) and neuter form in the complementiser, which could explain why the supposedly derived relative had an inflection, as well. On the other hand, as we know, the Middle English relative *þat* did not, probably because of its derivation from an invariable complementiser such as *þæt* itself.

As for the other West Germanic languages, just like in Old English, the Old Frisian complementiser *thet* and the homographic relative correspond to the nominative neuter singular form of the demonstrative *thi*, *thiu*, *thet*. The same can also be said for the Old Low Franconian demonstrative *thie*, *that* and the conjunction *that*, as well as for the Old Saxon *that*, which was used as the only form from the beginning.

4.4.3. North Germanic

As we can see from Table 17, Old Norse argumental complementisers and consecutive conjunctions were generally expressed by the conjunction *at*, which could, however, also be compounded with an adverb to specify its meaning (e.g. the adverb *svát* gives the subordinate a consecutive meaning).

Among the instances included in the corpus, the sole exception is the use of the relative particle *sem*, but Krause and Slocum (2005) do mention that the relative particles (especially *er*) and the conjunction *at* did at times overlap in their function. However, the interchangeability between the two functions and the relative forms could be just the other side of the same phenomenon which we have found in Gothic, as well as in some of the early West Germanic languages, i.e. the affinity between relative and argumental complementisers. While in Old Norse, in fact, the two functions could be expressed alternatively by *at*, *er* or *sem*, in the other languages instantiated the same element could be used to render both functions.

	Inflection	POS	Argumental Complementiser		Consecutive Conjunction		Tot.
			No.	Texts, Dating	No.	Texts, Dating	
<i>at</i>	UNINFLECTED	Conjunction	48	<i>Battle of Stamford Bridge</i> (y. 1066); <i>Íslendingabók</i> (1067-1148); <i>Edda</i> (1179-1241); <i>Brennu-Njáls Saga</i> (1250-1275); <i>Poetic Edda</i> , 1270s; <i>Egils Saga</i> , late 14 th cent.; <i>Grettis Saga</i> , 14 th cent.; <i>Hrólfs saga</i> , late 14 th cent.	26	<i>Battle of Stamford Bridge</i> (y. 1066); <i>Edda</i> (1179-1241); <i>Poetic Edda</i> , 1270s; <i>Egils Saga</i> , late 14 th cent.; <i>Grettis Saga</i> , 14 th cent.; <i>Hrólfs saga</i> , late 14 th cent.	74
<i>svát</i>		Adverb (<i>svá</i>) + Conjunction (<i>at</i>)	0	-	2	<i>Poetic Edda</i> , 1270s	2
<i>sem</i> ⁴⁰	UNINFLECTED	Relative Particle	±	<i>Hrólfs saga</i> , late 14 th cent.	0	-	±
Tot.			48	28			76

Table 18. Distribution of Old Norse conjunctions in the corpus.

As example (39) and (40) show, the conjunction *at* seems to be fully grammaticalised as both an argumental complementiser and, unlike Old English, also as a consecutive conjunction. In sentence (40), as in many other examples throughout the corpus, the conjunction *at* is, in fact, preceded by the adverb *svá*+ADJ/ADV and, at least based on the instances found in the corpus, this structure seems to have been far more common in Old Norse than in Old

⁴⁰ This form does not derive from the Proto-Indo-European *só-, *sá-, *tód and is therefore not relevant for the purposes of this study. This instance is not included in the sum total reported in the last row of Table 18.

English – thus probably implying a greater degree of grammaticalisation in the first.

- (39) OE Óláfr inn Digri gørði skýrt **at** hverr maðr skyldi
 Olaf the thick made definite ARG each man should
 gjalda konungi hálfu mörk.
 pay king.DAT.M.SG half mark
 ‘Olaf the Thick made definite that each man should pay half mark to the king.’

[cf. Appendix, no. 32]

- (40) OE [...] ok svá mikit gerðisk af því **at** þeir vildu
 and so far went with dem.DAT.M.PL CONJ they desired
 eigi nefna guð.
 no.longer mention god
 ‘[...] And went so far with this, that they no longer desired to mention God.’

[cf. Appendix, no. 33]

As already anticipated [cf. Section 2], however, it is argued that *at* might not really derive from the Proto-Indo-European *só-, *sá-, *tód – the reasons behind this uncertainty being the following:

- While in the West Germanic languages one single form (with very minor orthographic differences, e.g. the German *das/dass*) is used to express all the functions we are taking into account, in all the Present-Day Scandinavian languages, there is some formal difference between the language-specific demonstrative and complementiser;
- An homographic preposition and infinitive conjunction (meaning *to*, *by* and *near*, and deriving from the Proto-Indo-European preposition **ad*) exists in all the Present-Day Scandinavian languages – and it is argued that this could be the true origin of *at*;
- Assuming, as we previously have [cf. Section 2], that the relatives were grammaticalised directly from the demonstrative and thus before the argumental complementisers, it would be hardly acceptable to instantiate that the North Germanic languages had skipped (also according to the corpus data) the intermediate passage, i.e. the grammaticalisation of the relative *d-Pronouns*, but then featured that of the complementisers.

This notwithstanding, from a phonetic point of view, the derivation of *at* from the Proto-Indo-European **só-*, **sá-*, **tód*, and then especially from the Proto-Germanic **pat*, would be perfectly acceptable if we hypothesise – as already discussed [cf. Section 2] – the loss of the initial consonant.

Furthermore, from a typological perspective, while we have proof from the West Germanic languages of complementisers deriving from demonstratives, we have no such evidence as far as the derivation of complementisers from the Proto-Indo-European preposition **ad* is concerned. In Old English, for instance, the reflex of **ad* has preserved its original meaning until today in the form of the preposition *at*.

In conclusion, should the argumental complementiser have indeed grammaticalised before the relative, the third reason for doubting of the origin of the Old Norse *at* would be easily discredited: it would, in fact, be more plausible to believe that the North Germanic languages might have missed the final step of the grammaticalisation process, rather than thinking they might have just skipped a previous one.

On that note, it is also worth stressing that, as already discussed, according to Vigfusson (1874:29), the archaic Icelandic argumental complementiser *ad* (i.e. the reflex of the Old Norse *at* from which the Present-Day Icelandic *að* derives) could also be used as a relative. This fact could either be a further proof of the affinity between the relative and the complementiser (as seen, for instance, between the Old Norse *er* and *at*) or an indicator that, at some point in the history of the Scandinavian languages, complementisers did grammaticalise into relatives – thus generating a competing form with the already established relative particle *er*. Either way, however, based on the evidence we have from Old Norse and the Present-Day North Germanic languages, we can maintain that, despite the initial instability of the two forms, the particle *er* and the conjunction *at* have prevailed in the end, respectively as relative marker and argumental complementiser.

5. Discussion

In an attempt to make sense of the results described in the previous section of this paper, this chapter will set the evidence gathered from the linguistic corpus against the existing literature presented in Section 2. In doing so, this

study will put forward some hypotheses that aim to explain the discrepancies between the empirical data and other scholars' accounts.

Taking into consideration both the earlier phases and the Present-Day stage of the Germanic languages and based on the analysis carried out in the previous chapter [cf. Section 4.1], we can conclude that, although the determiner function of the demonstrative developed at different times in each individual language (especially interesting is the already mentioned late grammaticalisation of determiners in Old Low Franconian and their incomplete establishment in Old Norse), nowadays they all feature demonstrative forms which can be used equally as pronouns or as adjectives.

Furthermore, we know that the use of the demonstrative as a third-person pronoun has been largely dropped over time and we can assume that Old English was one of the first Germanic languages to stop employing it as such – provided that it had ever been an established use at all.

While the distribution of demonstratives was still significantly unstable in the earliest stages of most Germanic languages (as already seen, for instance, the Old Norse demonstrative could be expressed through several competing forms, with slight differences in meaning), the origin of at least one of their demonstratives from the anaphoric Proto-Indo-European *só-, *sá-, *tód is rather uncontroversial and, from the point of view of historical linguistics, the explanation of its evolution into a demonstrative determiner when accompanied by a noun [cf. Section 2] seems relatively unproblematic.

As for articles [cf. Section 4.2], based on the analysis of corpus data and on the contributions made by other scholars, like Crisma (2020) and Stroh-Wollin (2020), we may maintain that all three branches of the Germanic languages did undergo a grammaticalisation process which led their language-specific reflex of a Proto-Indo-European demonstrative (be it *só-, *sá-, *tód, or a combination of other demonstrative stems, e.g. the Old Norse *inn*, *in*, *it* from the Proto-Indo-European *ko-+*eno-) to become a definite article. On these grounds we can also claim that, albeit at different moments in time, Gothic, Old English and Old Norse all became languages with grammaticalised definiteness.

Although the definite articles systems of the Present-Day Germanic languages are fairly heterogeneous, even within the same branch (e.g. the Ger-

man “near-article” vs. the English and Dutch “discrete articles”), we can now maintain that both the West Germanic free-standing articles and the Scandinavian definite suffixes all originate from a deictic element.

With regards to the relative function, on the other hand, it appears that, according to the corpus data, no early Germanic language (except for Old Frisian in the fourteenth century) featured relative pronouns which had been directly grammaticalised from any of the reflexes of the demonstrative *só-, *sá-, *tód⁴¹ [cf. Section 4.3].

On these grounds, we might deduce that, this grammaticalisation path must have occurred at a later stage and only in some languages – signally in those in which the demonstrative is today also a *d-Pronomen*. Nonetheless, as already seen, Vigfusson’s account (1874) leaves the question open as to whether the same patterns were undertaken by the Scandinavian languages, as well – thus implying that their *d-Pronomen* has gone lost over time.

Therefore, though *that* and its cognates do undoubtedly feature this function today (at least in the West Germanic languages) the origin and development, as well as the nature of the relative function, raise quite a few questions: was the relative originally a demonstrative? how and why did it become a relative? can the *that*-relative be described as a true relative pronoun, from a syntactic point of view? etc.

Although some of these questions have already been answered by other scholars [cf. Section 2], this study attempts to give its own corpus-based interpretation, while aiming to account for the Present-Day distribution and status of relatives across the Germanic languages.

As far as the origin of Proto-Indo-European relatives is concerned, several scenarios have been hypothesised, including the possibility that Proto-Indo-European did not have relative clauses at all and that any relative marker featured in the daughter languages was thus an innovation introduced after their split from the mother language. This notwithstanding, there is general agreement on the stems from which Indo-European relatives might have originated – either already in Proto-Indo-European, or at a later stage – and

⁴¹ Old English features a relative particle whose form might have derived from the demonstrative [cf. Section 4.3.2]

these are the anaphoric **i-* and the interrogative-indefinite **k^wó-/*k^wí-*. While some scholars believe that one of the two derived forms was the original Proto-Indo-European relative pronoun and the other was an innovation independently made by some languages, Clackson (2007) maintains that both co-existed in Proto-Indo-European and were in complementary distribution with one another (Cotticelli-Kurras 2020).

As Quiles and López-Menchero (2011) argue, the Proto-Indo-European relative pronoun **ios, *iā, *iod* (from the relative stem **jó-*) gave origin to – among others – the Ancient Greek relative pronoun *ὄς, ἦ, ὅ* and the Proto-Germanic concessive/conditional conjunction **iaba, *iabai* (although it clearly did not retain the relative meaning), and was represented in every early Indo-European language except for Proto-Italic and possibly Anatolian⁴², where it had probably gone lost (Quiles and López-Menchero 2011)⁴³. It should also be noticed, however, that Hellenist scholars such as Monteil (1963) claim that, beside the relative pronoun deriving from **jó-*, Homeric Greek could also express the relative function by means of a pronoun deriving from the demonstrative **só-, *sá-, *tód*⁴⁴, possibly just like the Germanic reflexes of **tód* can be used as relative pronouns, as well.

As for the second relative pronoun, it derived, as already mentioned, from the interrogative-indefinite pronouns **qis, *qid* (substantive form) and **qos, *qā, *qod* (adjective form) – and, ultimately, from the Proto-Indo-European **kwo/i-* stem. These forms, in turn, originated interrogative and indefinite pronouns such as the Greek *τίς, τί* (and the clitic *τις, τι* for the indefinite) and the Latin *quis, quid*, but also the Proto-Germanic root **h^wa-*, from which all the Present-Day Germanic interrogative *w-Pronomen* stem. Interestingly,

⁴² The origin of the Hittite *ya-* from the Proto-Indo-European **jo-* is debated.

⁴³ Since it was present in almost every early Indo-European language, Quiles and López-Menchero (2011) believe that **ios, *iā, *iod* was indeed one of the two original Proto-Indo-European relative pronouns. It is more plausible to assume that it went lost in individual branches (i.e. Proto-Italian and possibly Anatolian) than to believe that all the other languages made the same innovation independently from one another.

⁴⁴ Over time, however, the relative stemming from **jó-* prevailed and became standard in Classic Greek.

however, whereas the Greek *τίς*, *τί* maintained the indefinite semantic and did not acquire a relative function, the Latin *quis*, *quid* did at some point grammaticalise into a relative pronoun, through the *diptyque normal* construction.

Assuming, thus, that, as De Roberto (2010) claimed, Latin only introduced relatives through the grammaticalisation of its interrogative-indefinite pronoun⁴⁵, we should be able to instantiate an earlier stage, before the grammaticalisation occurred [cf. example (5) in Section 2]. In this example (41a) we can still see, in fact, the personal pronoun *is* anaphorically referring to the interrogative-indefinite pronoun *quis*.

- (41) a. LAT ***Quis** *bene amat, is bene castigat.*
 IND.M.SG well loves, PERS.M.3SG well chastises
 ‘Someone well loves. He well chastises.’
 b. LAT **Qui** *bene amat, bene castigat.*
 REL.M.SG well loves, well chastises
 ‘Who well loves, well chastises’
 c. EN ***That** *loves well, well chastises.*
 d. EN **Who(ever)** *loves well, well chastises.*

We can thus assume the following step to have been the one described by De Roberto (2010), in which the anaphoric element is absorbed by the interrogative-indefinite pronoun, which thus acquires a new form (from *quis*, *quid* to the relative *qui*, *quae*, *quod*) and becomes a relative conjunction with an optional [+indefiniteness] feature (41b).

Following on this example, and in the light of what has been discovered so far, however, it could prove useful to address the grammaticalisation of Germanic relatives and their distribution by putting forward a different hypothesis, which is partly contrasting with the account given in the literature review [cf. Section 2].

As we know, besides the *w-Pronomen* deriving from the interrogative-indefinite pronoun **k^wó-* (> **hwo-* > *wh-*), the Present-Day West Germanic languages also feature *d-Pronomen*, i.e. relatives which derive from a Pro-

⁴⁵ Assuming that Proto-Italic had indeed lost the relative deriving from **ios*, **iā*, **iod*, it is likely that Latin initially lacked relatives.

to-Germanic demonstrative and, ultimately, from the Proto-Indo-European anaphoric pronoun *só-, *sá-, *tód. Remarkably, the same competition we can find today among the West Germanic relatives (i.e. between *w-Pronomen* and *d-Pronomen*) is supposed to have occurred in Proto-Indo-European, as well. As we have seen, in fact, Proto-Indo-European is assumed to have featured both an attributive-restrictive relative originating from the interrogative-indefinite pronoun and an appositive-explicative relative derivating from a demonstrative (Quiles and López-Menchero 2011).

While in Proto-Indo-European the two competing relatives seem to have been in a complementary distribution, however, as we have seen in Section 2, the status of the relatives in the Present-Day Germanic languages appears to be somewhat more complicated. For the moment, though, let us only consider that, in English, *w-Pronomen* may introduce either free or headed relative clauses, while *d-Pronomen* may only introduce headed ones.

On these grounds, if we were now to translate the previous sentence in English, we would have but one choice: the only grammatically acceptable way to preserve the *diptyque normal* structure is, in fact, interpreting the first clause as a free relative (36d) and thus render the Latin *qui* with the English *w-Pronoun who* (which has indeed an [+indefiniteness] feature).

Therefore, since the use of *d-Pronomen* in free relative clauses would be ungrammatical (41c), we could argue that, contrarily to previous assumptions deriving from the existing literature, the *diptyque normal* construction was not related to the grammaticalisation of demonstrative pronouns into relative *d-Pronomen*, but rather to the grammaticalisation of the Proto-Germanic (among other languages, e.g. Latin and Hittite) interrogative-indefinite pronoun (deriving from *h^wa-) into a relative *w-Pronomen*.

Furthermore, this interpretation would be typologically more acceptable, since this grammaticalisation path would be identical to the one which occurred in Latin and the interrogative-indefinite pronoun would be the one prevailing over the anaphoric element in both the Italic and the Germanic branch.

On the other hand, based on the structure of the sentence, we could interpret the grammaticalisation of demonstrative pronouns into relative *d-Pronomen* as originating from the *diptyque inverse* construction and as mediated by the grammaticalisation of the argumental complementiser.

More specifically, let us take the grammaticalisation of the argumental complementiser as our starting point. In Latin (42a) [cf. example (6) in Section 2], we find a demonstrative at the end of the first clause and the already grammaticalised relative (through the *diptyque normal* construction) at the beginning of the second one, which absorbed the demonstrative and was grammaticalised into an argumental complementiser.

- (42) a. LAT **Legati renuntiaverunt hoc, quod Pompeium in potestate haberent.*
 envoys reported DEM.ACC.N.SG REL.ACC.N.SG Pompeius in power had.3 PL
 ‘The envoys reported this: that they had Pompeius in (their) power.’
 b. LAT *Legati renuntiaverunt quod Pompeium in potestate haberent.*
 envoys reported ARG Pompeius in power had.3PL
 ‘The envoys reported that they had Pompeius in (their) power.’
 c. EN **The envoys reported this, | that they had Pompeius in their power.*
 d. EN *The envoys reported | that they had Pompeius in their power.*

Conversely, according to the hypothesis that this study has just proposed, in the Germanic languages we would have two demonstratives (42c) – the second of which was supposedly the anaphoric element which, in the presence of specific categories of verbs like that of the *verba dicendi*, prevailed over the other demonstrative and was grammaticalised into an argumental complementiser (42d).

However, if we were to compare an argumental subordinate (42c) with a relative one featuring a *d-Pronomen* (43a), we would obtain two surface-identical structures – the only difference being that the relative pronoun is always preceded by its antecedent, while the argumental clause constitutes the argument of the predicate.

- (43) a. EN *I have read this (book) | that you gave me.*
 b. DE *Ich habe das Buch gelesen, das du mir gegeben hast.*
 I have ART.ACC.N.SG book read.PP DEM.ACC.N.SG you to.o.me given have
 ‘I have read the book that you have given me.’
 c. NL *Ik heb het boek gelezen dat je me had gegeven.*
 I have ART.N.SG book read.PP DEM.N.SG you to.me have given.PP
 ‘I have read the book that you have given me.’

The striking similarity between the two constructions hints at their possible common origin and it is thus believable that the relative *d-Pronomen* could simply be an argumental complementiser which has been reanalysed when in the absence of specific categories of verbs (like *verba dicendi*, *sentendi*, *iubendi*, etc.) and whose meaning has thus been extended.

Furthermore, although the structure of sentence (43a) seems to be modelled on the *diptyque inverse* construction (and contrarily to what has been previously outlined in Section 2), from this perspective, it would no longer be necessary to assume the absorption of an antecedent by the *d-Pronomen* in order to explain its reanalysis into a subordinator. According to this hypothesis, in fact, the *d-Pronomen* would merely be an extension of the argumental complementiser and would therefore have been a subordinator from the beginning.

Therefore, while rejecting Axel-Tober's account, this interpretation revisits the Neogrammarians' theory – according to which the argumental complementiser had grammaticalised directly from the demonstrative [cf. Section 2]. Contrarily to the traditional account however, this study proposes that the demonstrative from which the complementiser came had maintained its anaphoric nature and had thus always belonged to the second clause (42c).

As for the other main West Germanic languages, i.e. German (43b) and Dutch (43c), they seem to have taken the same grammaticalisation path.

From a cross-linguistic point of view, the *diptyque normal* thus appears to have originated relatives heading free relative clauses (i.e. indefinite relative pronouns or *w-Pronomen*), while the *diptyque inverse* construction has arguably been the model structure for the grammaticalisation of argumental complementisers and relatives introducing headed relative clauses (i.e. *d-Pronomen*).

While wondering why this only applies to West Germanic languages and not to the Scandinavian ones might lead us into analysing a false problem (the North Germanic languages – except for Danish – only use relative particles and do not have any *w-Pronomen* either), the grammaticalisation of the “Latin *d-Pronomen*” (i.e. the relative pronoun introducing a headed relative clause) is surely worth discussing.

Based on example (44), in fact, they seem to have followed the same grammaticalisation pattern we have described for the West Germanic *d-Pronomen*, in that the structure of the argumental subordinate (42b) seems to be identical to the relative one reported below.

- (44) LAT *Magistra hoc laudat quod discipulus fecit.*
 Teacher DEM.ACC.N.SG praises REL.ACC.N.SG student did
 ‘The teacher praises the thing (=this) which the student did.’

Unlike the West Germanic languages, however, the “Latin *d-Pronomen*” maintained the same surface-identical form as the indefinite relative pronoun probably because the argumental complementiser from which it came had, in turn, originated from the indefinite-relative itself.

Although we cannot be sure as to what the reason for the indefinite relative to grammaticalise into an argumental complementiser was, and we do not know for definite why the Germanic complementiser came, instead, from the demonstrative, it seems acceptable to conclude that the Germanic *d-Pronomen* have taken the form of the demonstrative because they came from the argumental complementiser, which had originated from the demonstrative, in the first place.

After all, as we have already seen [cf. Section 4.4.1] shared features between relative and argumental clauses could be found even in Gothic, as both clauses are headed by conjunctions or pronouns containing the relative marker *ei*. While, in fact, just like in Latin, neither form derives from the grammaticalisation of the demonstrative, we may interpret the similarity between relative and argumental heads as a typological feature, which is independent of the ultimate derivation of either elements.

As for the argumental complementiser itself, based on the corpus data analysed in the previous chapter, it seems to have grammaticalised directly from the demonstrative derived from the Proto-Indo-European **só-*, **sá-*, **tód* in all of the languages taken in consideration, except for Gothic – which, however, separated from Proto-Germanic quite early and did not have any significant contact with the other Germanic languages afterwards.

On the other hand, the fact that this grammaticalisation path did occur in the other Germanic languages at a stage when the *d-Pronouns* were not yet

grammaticalised seems to further confirm the hypothesis that the grammaticalisation process of relatives was a later extension of that of the argumental complementisers.

Furthermore, also thanks to typological evidence, we may be convinced that the Old Norse *at* did in fact derive from the Proto-Indo-European **só-*, **sá-*, **tód* [cf. Section 4.4.3].

Finally, as far as consecutive conjunctions are concerned, their grammaticalisation process, too, is considered to be another extension of that of the argumental complementisers. As such, however, the grammaticalisation degree of the conjunction used to head consecutive clauses seems to differ from language to language. In almost all of the Old English entries, in fact, the supposed consecutive clauses do lack their signature structure, i.e. *so+ADJ/ADV... that* – which, on the other hand seems to have already been quite common in Old Norse.

6. Conclusion

This study has attempted to examine the existing literature [cf. Section 2] in the light of the data that have been collected from the early stages of the main Germanic languages and gathered into a linguistic corpus [cf. Section 3, Appendix].

As it is evident from what has been said so far in this paper, an in-depth analysis of the corpus has often led towards new interpretations which are at least partly contrasting with the initial account described in the literature review. At the same time, the entire process of putting forward some hypotheses and trying to verify them, has given rise to new questions, which could represent possible future developments of this research.

Lang.	Demonstrative Pronouns		Demonstrative Determiners		Articles		Relatives		Conjunctions			Tot.
									ARG	CONS		
GOT	<i>sa, so, þata</i>	13	<i>sa, so, þata</i>	8	<i>sa, so, þata</i>	52	-	0	-	0	0	73
OE	<i>sē, sēo, ðæt</i>	13	<i>sē, sēo, ðæt</i>	12	<i>sē, sēo, ðæt</i>	11	<i>þe(?)</i>	3‡	<i>þæt</i>	17	1	54
OHG	<i>daz, thaz</i>	1	<i>daz, thaz</i>	0	<i>daz, thaz</i>	0	-	0	<i>daz, thaz</i>	1	0	2

OS	<i>that</i>	0	<i>that</i>	1	<i>that</i>	0	-	0	<i>that</i>	0	0	1
OF	<i>thi, thiu, thet</i>	2	<i>thi, thiu, thet</i>	0	<i>thi, thiu, thet</i>	2	<i>thi, thiu, thet</i>	1	<i>thet</i>	1	0	6
OLF	<i>thie</i>	1	-	0	-	0	-	0	<i>that</i>	0	0	1
ON	<i>sá, sú, þat</i>	83	<i>sá, sú, þat</i>	11	<i>sá, sú, þat</i>	13	-	0	<i>at</i>	48	28	183
Tot.		113		32		78		1		67	29	320

Table 19. Distribution of relevant forms per function in the corpus.⁴⁶

Therefore, based on a careful study of the linguistic corpus [Table 19] and of the subject matter, we could tentatively describe the grammaticalisation paths of the Proto-Indo-European **só-*, **sá-*, **tód* across the Germanic languages as follows.

In agreement with the existing literature, we can maintain that the anaphoric demonstrative pronoun **só-*, **sá-*, **tód* became a demonstrative determiner in the earliest stages of most Germanic languages through a grammaticalisation process that allowed for it to be followed by a noun, despite its being a pronoun. Although each language probably developed this function at different times, based on the entries found in the corpus, we cannot find any such instance in Old Low Franconian and the representation of the determiner in Old Norse seems to have been unstable [cf. Section 4.1].

In each individual Germanic language, then, the demonstrative determiner deriving from **só-*, **sá-*, **tód* lost its [-deictic] feature and acquired the [+definiteness] one, thus beginning to function as a definiteness marker (which generally maintained the same form as the determiner). In Old Norse, however, this same process also involved a Proto-Scandinavian demonstrative pronoun which later produced Present-Day Scandinavian definite suffixes [cf. Section 4.2].

As for the following steps, this study distances itself from what has been

⁴⁶ This table provides an overview of the occurrences per grammaticalised function of the language-specific forms deriving from **só-*, **sá-*, **tód*. This table only includes the instances whose indicated function was confirmed by the corpus analysis described in Section 4. Once again, Gothic, Old English, and Old Norse are the most represented languages in this table, as, due to a specific methodological decision [cf. Section 1 and 3], they are the most represented languages in the corpus.

suggested by other scholars by proposing that the grammaticalisation of argumental complementisers might have taken place before that of the relatives. Arguably, in fact, in all of the Germanic languages taken into consideration – except for Gothic – the demonstrative **só-*, **sá-*, **tód* grammaticalised into an argumental complementiser through a *diphthyc inverse* construction [cf. Section 4.4, Section 5].

Furthermore, based on the evidence provided in Section 4.4.3, this paper supports the hypothesis that the Old Norse *at* and its Present-Day descendants [Table 20] are in fact all reflexes of the anaphoric **tód*.

According to this study, two more functions have then originated from the complementiser, through a *diphthyc inverse* structure, i.e. the relative marker function and that of the consecutive conjunction. These two grammaticalisation paths – interpretable as independent patterns of extension of the meaning of the complementiser – could be described as later processes which had generally not been established (especially as far as the relatives are concerned) in the time span covered by the linguistic corpus compiled for this study [cf. Section 4.3, Section 4.4].

Finally, with regard to the origin of relatives, this research claims that, contrarily to other existing accounts, the *diphthyc normal* structure only gave rise to the relative *w-Pronomen* – i.e. the type of relative pronoun which normally heads free relatives – while, as already mentioned, the relative *d-Pronomen* – which is normally used for headed relatives – derived from the extension of the meaning of the complementiser in the presence of an antecedent of the relative itself.

FUNCTION	EN	DE	NL	DA	NO	SV	IS
Demonstrative Pronoun/ Determiner	<i>that</i> < <i>*tód</i>	<i>das</i> < <i>*tód</i>	<i>dat</i> < <i>*tód</i>	<i>det</i> < <i>*tód</i>	<i>det</i> < <i>*tód</i>	<i>det</i> < <i>*tód</i>	<i>það</i> < <i>*tód</i>
Definite article	<i>the</i> < <i>*tód</i>	<i>das</i> < <i>*tód</i>	<i>het</i> ⁴⁷ < <i>*tód</i>	<i>det</i> < <i>*tód</i> +suf- fi<DEM	suffix <DEM	<i>det</i> < <i>*tód</i> +suf- fi<DEM	suffix <DEM

⁴⁷ Cf. footnote no. 25 in [Section 4.2.2].

<u>Conjunction</u> (ARG, CONS)	<i>(that),</i> <i>so that</i> <*tód	<i>(dass),</i> <i>sodass</i> <*tód	<i>(dat),</i> <i>om-</i> <i>dat/</i> <i>zodat</i> <*tód	<i>(at),</i> <i>så at</i> <*tód	<i>(at),</i> <i>Slik</i> <i>at/</i> <i>Så at</i> <*tód	<i>(att),</i> <i>Så att</i> <*tód	<i>(að),</i> <i>svo að</i> <*tód
Relative	<i>that</i> <*tód	<i>das</i> <*tód	<i>dat</i> <*tód	RELA- TIVE PARTI- CLE	RELA- TIVE PAR- TICLE	RELA- TIVE PARTI- CLE	RELA- TIVE PAR- TICLE

Table 20. Present-Day relevant forms per function and derivation.⁴⁸

Abbreviations

ACC	Accusative	ME	Middle English
ADJ	Adjective	N	Neuter
ADV	Adverb	NL	(Present-Day) Dutch
ARG	Argumental Complementiser	NO	(Present-Day) Norwegian
A-S	Anglo-Saxon	NOM	Nominative
CONJ	Conjunction	OE	Old English
CONS	Consecutive Conjunction	OF	Old Frisian
DA	(Present-Day) Danish	OHG	Old High German
DAT	Dative	OLF	Old Low Franconian
DE	(Present-Day) German	ON	Old Norse
DEM	Demonstrative Pronoun	OS	Old Saxon
DET	Determiner	PERS	Personal Pronoun
EN	(Present-Day) English	PG	Proto-Germanic
F	Feminine	PIE	Proto-Indo-European
GEN	Genitive	PL	Plural
GR	Ancient Greek	PP	Past Participle
IND	Indefinite Pronoun	PRSP	Present Participle
INF	Infinitive Conjunction	REL	Relative (Pronoun/Particle)
INSTR	Instrumental	SG	Singular
IS	(Present-Day) Icelandic	SV	(Present-Day) Swedish
IT	(Present-Day) Italian	SG	Singular
LAT	Latin	SV	(Present-Day) Swedish
M	Masculine		

⁴⁸ This table provides an overview of the Present-Day reflexes of *tód and, where relevant, of the alternative forms used by the Scandinavian languages.

Bibliography

Primary Sources

- (A) “RUNES: Forschungsprojekt der Akademie der Wissenschaften zu Göttingen”, 2021, *RuneS*.
- (B) Krause, Todd B. & Slocum, Jonathan, 2005, “Early Indo-European Online: Old Norse Online”, *The University of Texas at Austin - Linguistics Research Center*.
- (C) Krause, Todd B. & Slocum, Jonathan, 2006, “Early Indo-European Online: Gothic Online”, *The University of Texas at Austin - Linguistics Research Center*.
- (D) Robinson, Orrin W., 1992, *Old English And Its Closest Relatives*, Stanford, California, Stanford University Press.
- (E) Slocum, Jonathan & Lehmann, Winfred P., 2007, “Early Indo-European Online: Old English Online”. *The University of Texas at Austin - Linguistics Research Center*.

References

- Allen, Cynthia L., 2019, “The Definite Article In Old English: Evidence From Ælfric’S Grammar”, in Yáñez-Bouza N., Moore E., van Bergen, L. & Hollman, W. B. (eds.). *Categories, Constructions And Change In English Syntax*, Cambridge University Press, 130-146.
- Axel-Tober, Katrin, 2017, “The Development of the Declarative Complementizer in German”, *Language* 93(2), e29-e65.
- Bacskai-Atkari, Julia, 2020, “English Relative Clauses In A Cross-Germanic Perspective”, *Nordlyd* 44(1), 93-115.
- Biberauer, Theresa & Roberts, Ian, 2017, “Parameter Setting”, *The Cambridge Handbook of Hystorical Syntax*, 134-162.
- Bisang, Walter, & Malchukov, Andrej, 2020, *Volume 1: Grammaticalization Scenarios From Europe And Asia*, Berlin, Walter de Gruyter & Co.
- Bybee, Joan L., 2006, “Language Change and Universals”, In Mairal, R. (ed), *Linguistic Universals*, Cambridge, Cambridge University Press, 179–194.
- Campbell, Lyle, 1998, *Historical Linguistics – An Introduction*, Cambridge, Massachussets, The MIT Press.
- Chantraine, Pierre, 1958, *Grammaire Homérique – Tome II: Syntaxe*, Paris, Librairie C. Klincksiek, 165.
- Chomsky, Noam, 1988, *Current Issues in Linguistic Theory*, Berlin, De Gruyter, 14-15.
- Clackson, James, 2007, *Indo-European Linguistics: An Introduction*, Cambridge, Cambridge University Press.

- Cotticelli-Kurras, Paola, 2020, "Relative Clauses And Wh-Words In Ancient IE Languages: Etymology, Function, Distribution, Classification", Lecture, Università degli Studi di Verona.
- Crisma, Paola, 2020, *Language Contact And The Early English Noun Phrase*, Canterano (RM), Aracne editrice.
- Dahl, Östen, 2018, "Grammaticalization in the Languages of Europe", In In Narrog, H. & Heine, B. (eds.), *Grammaticalization From A Typological Perspective*, Oxford, Oxford University Press.
- De Roberto, Elisa, 2010, "Correlative, Strutture", *Enciclopedia Dell'italiano*, Treccani.
- Dunkel, George, 2014, *Lexikon der indogermanischen Partikeln und Pronominalstämme*, Heidelberg, Universitätsverlag Winter.
- Fontana, Alessandro, 2022, *The Proto-Indo-European Anaphoric and Relative Pronouns: Analysis and Considerations*.
- Fuß, Eric & Grewendorf, Günther, 2014, "Freie Relativsätze Mit D-Pronomen", *Zeitschrift Für Sprachwissenschaft* 33 (2) 165-214.
- Harbert, Wayne Eugene, 2007, *The Germanic Languages*, Cambridge, Cambridge University Press.
- Haspelmath, Martin, 2001, "The European Linguistic Area: Standard Average European". *Language Typology And Language Universals - Volume 2*, Berlin/New York, Walter De Gruyter.
- Haudry, Jean, 1979, "L'antéposition de la Relative en Indo-européen", *La Linguistique* 15 (1), 101-10.
- Heine, Bernd & Kuteva, Tania, 2010, "Grammaticalisation Theory as a Tool for Reconstructing Language Evolution", *The Oxford Handbook Of Language Evolution*, 512-527.
- Hendery, Rachel, 2012, *Relative Clauses in Time and Space - A Case Study in the Methods of Diachronic Typology*, Amsterdam/Philadelphia, John Benjamins Publishing Company.
- Keretchashvili, George, 2018, "Pronoun Paradigms". *Chrestomathy Of Gothic And Anglo-Saxon Written Records*.
- Krause, Todd B., & Slocum, Jonathan, 2005, "Early Indo-European Online: Old Norse Online", *The University of Texas at Austin - Linguistics Research Center*.
- Krause, Todd B., & Slocum, Jonathan, 2006, "Early Indo-European Online: Gothic Online", *The University of Texas at Austin - Linguistics Research Center*.
- Kroch, Anthony, 2001, Syntactic Change, In Baltin, M. & Collins, C. (eds.), *Handbook of Syntax*, Blackwell.
- Lee, Soojin, 2006, "That Or Which?: The That's That Of Which Is Which".

- Lehmann, Christian, 2002, "Thoughts on Grammaticalisation", *Arbeitspapiere des Seminars für Sprachwissenschaft der Universität Erfurt*, Volume 9.
- Lyons, Christopher, 1999, *Definiteness*, Cambridge, Cambridge University Press.
- Minard, Armand, 1936, *La Subordination dans la Prose Védique*, Paris, Les Belles Lettres.
- Monteil, Pierre, 1963, *La phrase relative en grec ancien*, Paris, Klincksieck.
- Narrog, Heiko & Heine, Bernd, 2018, "Introduction: Typology and Grammaticalisation", In Narrog, H. & Heine, B. (eds.), *Grammaticalization From A Typological Perspective*, Oxford, Oxford University Press.
- Pat-El, Na'ama, 2021, "Typological Approaches And Historical Linguistics", *The Handbook Of Historical Linguistics – Volume II*, John Wiley & Sons, Inc., 183-195.
- Quiles, Carlos & López-Menchero, Fernando, 2011, *A Grammar Of Modern Indo-European*, Badajoz: Indo-European Language Association.
- Radford, Andrew, 2004, *Minimalist Syntax*, Cambridge, Cambridge University Press, 9.
- Roberts, Ian & Roussou, Anna, 2007, "Syntactic Change – A Minimalist Approach to Grammaticalization", *Beiträge zur Geschichte der deutschen Sprache und Literatur (PBB)* 129 (3), 469-476.
- Robinson, Orrin W., 1992, *Old English And Its Closest Relatives*. Stanford, California, Stanford University Press.
- Rydberg-Cox, Jeff, 2021, *A Digital Tutorial For Ancient Greek*, Accessed 23/09/2021.
- Sijs, Noline van der, 2010, "Dat (Voornaamwoord, Voegwoord)". *Etymologiebank.nl*, Instituut voor de Nederlandse Taal.
- Skrzypek, Dominika, Piotrowska, Alicja & Jaworski, Rafał, n.d., *The Diachrony Of Definiteness In North Germanic*, Leiden, Brill.
- Stroh-Wollin, Ulla, 2020, "Hinn And Hinn: Early Icelandic As The Clue To The History And Etymology Of Two Old Scandinavian Words", *Nordic Journal Of Linguistics* 43 (2), 205-228.
- Velupillai, Viveka, 2012, *An Introduction To Linguistic Typology*, John Benjamins Publishing Company.
- Vigfusson, Gudbrand, 1874, "AT", *An Icelandic-English Dictionary*, Oxford, The Clarendon Press.
- Viti, Carlotta, 2013, "Forms and Functions of Subordination in Indo-European", *Historische Sprachforschung / Historical Linguistics* 126, 89-117.
- Viti, Carlotta, 2015, *Variation und Wandel in der Syntax der alten indogermanischen Sprachen*, Tübinger Beiträge zur Linguistik (TBL).
- Zimmermann, Richard, 2012, "Grammar Competition In Old English Relative Clauses", Lecture, Université de Genève.

Appendix – Corpus selection

Key:

No = Progressive number of entry

L = Language

Or. = Origin

SF = Syntactic function

Inf. = Inflection

So. = Source [cf. Primary Sources in Bibliography]

bf. = before

NO	L	ENTRY	TEXT	OR.	SF	INF.	TRANSLATION	SO.
1	GOT	Warþ þan in dagans jainans , urrann gagrefts fram kaisara Agustau, gameljan allana midjungard.	Gospels (4 th cent.)	jáins	DET	ACC. M.PL	And it came to pass in those days, that there went out a decree from Caesar Augustus, that all the world should be taxed.	C
2	GOT	Warþ þan in dagans jainans, urrann gagrefts fram kaisara Agustau, gameljan allana midjungard. Soh þan gilstrameleins frumista warþ at [wisandin kindina Swriais] raginondin Saurim Kwreinaiau.	Gospels (4 th cent.)	sa, so, þata	DET	NOM. F.SG	And this taxing was first made when Cyrenius was governor of Syria.	C
3	GOT	sa ist sunus meins sa liuba, þamma hausjaip	Gospels (4 th cent.)	sa, so, þata	DEM	NOM. M.SG	This is my beloved Son: hear him.	C
4	GOT	jah þatei gasahv jag gahausida þata weitwodeip	Skeireins (5 th cent.)	sa, so, þata	DEM	ACC. N.SG	and what he saw and heard, that he testifies	C
5	OE	Næfde se here, Godes þonces, Angelcyn ealles forswiðe gebrocod, ac hie wæron micle swiþor gebrocede on þæm þrim gēarum mid cēapes cwilde ond monna;	Anglo-Saxon Chronicles (Year 897)	sē, sēo, ðæt	DET	DAT. PL	The enemy had not, by the mercy of God, entirely crushed the English altogether, but they were afflicted much more in those three years by pestilence of cattle and of men;	E
6	OE	ealles swiþost mid þæm þæt manige þāra sēlestena cynges þēna þe þær on londe wæron forðferdon on þæm þrym gēarum.	Anglo-Saxon Chronicles (Year 897)	sē, sēo, ðæt	PERS	NOM. N.SG	most of all among them many of the best of the king's thanes who were there in the land died within those three years.	E

7	ON	<i>En þar galzk stundum meira, en stundum minna, unz Óláfr inn Þigri gærði skýrt at hverr maðr skyldi gjalda konungi hálfu mörk, sá er færi á miðli Norvegs ok Íslands, nema konur eða þeir menn er hann næmi frá.</i>	<i>On the Settling of Iceland</i> (Ari Þorgilsson 1067-1148)	<i>hann, hon, þat</i>	DET	NOM. M.PL	Sometimes more were paid, sometimes less, until Olaf the Thick made definite that each man should pay the king a half mark, whoever would travel between Norway and Iceland, except women or those men whom he should exempt.	B
8	ON	<i>Fyrir þá sök</i>	<i>Snorra Edda</i> (Snorri Sturluson 1179-1241)	<i>sá, sú, þat</i>	DET	ACC. F.SG	For that reason	B
9	OLF	<i>ginathi in uuarheide sina uue sal thia suocan?</i>	<i>Psalm 60</i> (n.d.)	-	DEM	-	Who shall question his grace and truth?	D

Table 21. Syntactic function: Demonstrative

NO	L	ENTRY	TEXT	OR.	INF.	TRANSLATION	So.
10	ON	<i>watehalihinohor(na) hahaskapihapuligi</i>	<i>Older Fuþark</i> (2 nd -8 th cent.)	<i>hinn</i>	-	may the horn wet this stone. cut the aftermath. may the mowing lie.	A
11	GOT	<i>wasuh þan nelwa pasxa, so dulþs Iudaie.</i>	<i>Gospels</i> (4 th cent.)	<i>sa, sō, þata</i>	NOM. F.SG	was then nigh the passover, the feast of the Jews	C
12	OE	<i>Þā hē þæt þā sumre tīde dyde, þæt hē forlēt þæt hūs þæs gebēorscipes, and ūt wæs gongende tō nēata scypene, þāra heord him wæs þære nihte beboden,</i>	<i>Bede's Account of the Poet Caedmon</i> (671-735)	<i>sē, sēo, ðæt</i>	ACC. N.SG	Then one time he did this, so that he left the house of the feast and was going out to the cattle shed (their care was entrusted to him for the night).	E

13	ON	<i>Dá mællti Haraldr Ængla konongr viðr Norðmenn þá er með hanum váro, «Kenndo þér þenn hinn myckla meðr þæim blá kyrтли oc hin faghra hialm, er þer skaut sér af hestinum frem?</i>	Battle of Stamford Bridge (1066)	sá, sú, þat	DAT. M.SG	Then Harald, king of the Angles, spoke with the Norsemen who were with him: 'Do you know that noble man with the blue kirtle and impressive helmet, who launched himself off his horse?'	B
14	ME	<i>þe wrecce men of þe land</i>	<i>Peterborough Chronicle (1137)</i>	þe	-	the wretched men of the land	-
15	ME	<i>þe wrecce men of þe land</i>	<i>Peterborough Chronicle (1137)</i>	þe	-	the wretched men of the land	-

Table 22. Syntactic function: Article

NO	L	ENTRY	TEXT	OR.	INF.	TRANSLATION	SO.
16	GOT	<i>usgeisnodedun þan allai þai hausjandans is ana frodein jah andawaurdjam is.</i>	Gospels (4th cent.)	sa, so, þata	NOM. M.PL	And all that heard him were astonished at his understanding and answers.	C
17	GOT	<i>Iesu sokeiþ Nazoraiu þana ushramidan;</i>	Gospels (4th cent.)	sa, so, þata	ACC. M.SG.	Jesus of Nazareth which was crucified;	C
18	GOT	<i>unte jabai afletip mannam missadedins ize, afletip jah izwis atta izwar sa ufar himinam</i>	Gospels (4th cent.)	sa, so, þata	NOM. M.SG	For if ye forgive men their trespasses, your heavenly Father will also forgive you:	C
19	GOT	<i>jah qaþ im: izwis atgiban ist kunnan runa þiudangardjos gudis, iþ jainaim þaim uta in gajukom allata wairþiþ</i>	Gospels (4th cent.)	sa, so, þata	DAT. M.PL	And he said unto them: 'Unto you it is given to know the mystery of the kingdom of God, but unto them that (are) without in parables all these things are done'	C
20	GOT	<i>Atsaihvip armaion izwara ni taujan in andwairþja manne du sailuan im; aiþþau laun ni habaiþ fram attin izwaramma þamma in himinam.</i>	Gospels (4th cent.)	sa, so, þata	DAT. M.SG	Take heed that ye do not your alms before men, to be seen of them: otherwise ye have no reward of your Father which is in heaven.	C

21	GOT	<i>if þu þan biðjaís, gagg in heþjon þeina jah galukands haurðai þeinai bidei du attin þeinamma þamma in fulhsnja, jah atta þeins saei saiþviþ in fulhsnja, usgibiþ þus in bairhtein.</i>	Gospels (4 th cent.)	sa, so, þata	DAT. M.SG	But thou, when thou prayest, enter into thy closet, and when thou hast shut thy door, pray to thy Father which is in secret; and thy Father which seeth in secret shall reward thee openly.	C
22	GOT	<i>ei sijai so arma-hairtiþa þeina in fulhsnja, jah atta þeins saei saiþviþ in fulhsnja, usgibiþ þus in bairhtein.</i>	Gospels (4 th cent.)	sa+ei	NOM. M.SG	But thou, when thou prayest, enter into thy closet, and when thou hast shut thy door, pray to thy Father which is in secret; and thy Father which seeth in secret shall reward thee openly.	C
23	OE	<i>alle bûtan ānum, sē wæs þæs aldormonnes godsunu; ond hē his feorh generede, ond þeah hē wæs oft gewundad</i>	Anglo-Saxon Chronicles (Year 755)	sē, sēo, ðæt	NOM. M.SG	all but one, who was his nobleman's godson; and he saved his life, though he was much wounded.	E
24	OE	<i>Þonne is ān port on sūðewardum þām lande, þone man hætt Sciringes hēal.</i>	Voyages of Oththere and Wulfstan (bf. 899)	sē, sēo, ðæt	ACC. M.SG.	There is a port in the south of that land, which one calls Skiringssal.	E
25	OE	<i>Eal þæt his man āþer oððe ettan oððe erian mæg, þæt lið wið ðā sē;</i>	Voyages of Oththere and Wulfstan (bf. 899)	sē, sēo, ðæt	NOM. N.SG.	All that a man can either graze or plough extends alongside the sea;	E
26	OE	<i>Geworhton ðā Wedra lēode hlāw on hliðe, sē wæs hēah ond brād, wægliðendum wide gesýne,</i>	Beowulf (1000)	sē, sēo, ðæt	NOM. PL.	the people of the weders built upon the hill a mound, which was high and wide, visible to seafarers from afar,	E
27	OE	<i>and eall þæt to fæsle frea ælmihtig habban wolde under hrof gefor to heora ætgifan, swa him ælmihtig weroda drihten þurh his word abead.</i>	Genesis (1000)	sē, sēo, ðæt	ACC. N.SG.	and all that the Lord Almighty would have for progeny. And he went under the roof as their provider, as the Almighty, the Lord of hosts, bade him by His word.	E

28	OE	he wæs milde þam godum mannum þe god lufedon	Peterborough Chronicle (1087)	sē, sēo, ðæt	DAT. PL.	he was gentle with those good men that love god	E
29	OF	ther hi on eskriuin hede tha tian bodo, tha skolde hi lera tha Israheliska folke	Asegabook (1300)	-	-	where he had written the ten commandments, which he should teach to the israelite folk	D

Table 23. Syntactic function: Relative

NO	L	ENTRY	TEXT	OR.	SF	TRANSLATION	So.
30	OE	Þā hē þæt þā sumre tide dyde, þæt hē forlēt þæt hūs þæs gebēorscipes, and üt wæs gongende tō nēata scypene, þāra heord him wæs þære nihte beboden,	Bede's <u>Account of the Poet Caedmon</u> (671-735)	þæt	CONS	Then one time he did this, so that he left the house of the feast and was going out to the cattle shed (their care was entrusted to him for the night).	E
31	OE	For þon cnyssað nū heortan geþōhtas, þæt ic hēan strēamas, sealtýþa gelāc sylf cunnige;	The Seafarer (10 th cent.)	þæt	ARG	therefore it strikes now the thoughts of the heart, that i the humble streams, the tumult of sea waves myself should test	E
32	ON	En þar galzk stundum meira, en stundum minna, unz Óláfr inn Digri gærði skýrt at hvern maðr skyldi gjalda konungi hálfu mörk, sá er færi á miðli Norvegs ok Íslands, nema konur eða þeir menn er hann næmi frá.	On the Settling of Iceland (Ari Þorgilsson 1067-1148)	at	ARG	Sometimes more were paid, sometimes less, until Olaf the Thick <u>made definite</u> that each man should pay the king a half mark, whoever would travel between Norway and Iceland, except women or those men whom he should exempt.	B
33	ON	ok svá mikit gerðisk af því at þeir vildu eigi nefna guð.	Snorra Edda (Snorri Sturluson 1179-1241)	at	CONS	and this went <u>so far</u> that they no longer desired to speak the name of god.	B

Table 24. Syntactic function: Conjunction

ARTICOLI, NOTE, SAGGI

*Analisi linguistiche di testi arcaici, riflessioni su aspetti
e problemi linguistici del mondo antico, linee e momenti
di preistoria e protostoria linguistica*

LUISA CORONA, ROSALBA NODARI*

“Finisco da scrivere con le parole più belle”. Lessico e strategie linguistiche dell’oralità in *Lettere da una tarantata*¹

Abstract

L'articolo propone un'analisi quantitativa e qualitativa dei principali fenomeni linguistici contenuti nelle lettere scritte negli anni Sessanta da Anna, una donna semicolta affetta da tarantismo, all'antropologa Annabella Rossi. Analizzeremo dapprima l'evoluzione del lessico all'interno del corpus, con un focus sulla descrizione di fenomeni tipici della dimensione discorsiva che emergono dalla scrittura di Anna. In accordo con i "New Literacy Studies", mostreremo come l'intreccio tra dominio orale e scritto permetta ad Anna sia di scrivere testi, sia di svolgere attività come il mantenimento della relazione con Annabella. Infine, la ricerca mostrerà come alcune delle caratteristiche linguistiche e testuali di questo tipo di scritture, che deviano dalle norme di produzione di testi che hanno come modalità di ricezione la lettura, costituiscono veri e propri correlati funzionali della modalità parlata.

Parole chiave: scritture semicolte, lessico, oralità, strategie discorsive, pragmaticalizzazione

The paper offers a quantitative and qualitative analysis of the main linguistic phenomena contained in the letters written by Anna, a semiliterate woman affected by tarantism, to the anthropologist Annabella Rossi during the 1960s. The analysis will show the evolution of the lexicon within the corpus, with a focus on the description of some of the discursive dimensions that emerge in Anna's writing. In accordance with New Literacy Studies, we will demonstrate how the intertwining of oral and written domains allows Anna to both write texts and do things such as maintaining her relationship with Annabella. Finally, the research will display that some specific linguistic and textual features of semiliterate writings which appear to deviate from the norms of written texts, instead constitute functional correlates of the speech modality.

Keywords: semiliterate writings, lexicon, orality, discourse strategies, pragmaticalization

* Luisa Corona, Università degli Studi dell'Aquila, luisa.corona@univaq.it; Rosalba Nodari, Università degli Studi di Siena, rosalba.nodari@unisi.it.

¹ Per fini esclusivamente accademici, l'attribuzione è da intendersi nella modalità seguente. Luisa Corona ha responsabilità di 2.2, 3, 5, 5.1, 5.2 e 5.3, mentre Rosalba Nodari ha responsabilità di 2, 2.1, 4, 4.1 e 6; i parr. 1 e 5.4 sono stati scritti congiuntamente. Ringraziamo per il cortese supporto Domenico Ferraro della casa editrice Squilibri.

1. Introduzione

Lo scambio epistolare tra Anna, tarantata salentina, e l'antropologa Annabella Rossi ha uno statuto particolare nella storia della linguistica italiana. È grazie al coinvolgimento nello studio di questi testi di Tullio De Mauro da parte della stessa Annabella Rossi che prende avvio la teorizzazione dello studioso sull'italiano popolare unitario (cf. De Mauro 2005 [1970]¹). L'italiano di Anna, osservato nei suoi dettagli linguistici, fa emergere una serie di tratti diagnostici specifici che non sono ascrivibili esclusivamente alla variazione diatopica, diastratica o diafasica ma sono riscontrabili in altre scritture semicolte e paiono rispondere a una serie di regole. Non si tratta quindi di un italiano malformato, bensì della risposta a un bisogno comunicativo che prende vita attraverso il racconto che la donna fa di sé.

Purtuttavia, il rapporto instaurantesi tra testo e racconto non sembra essere stato fatto oggetto di analisi. Le lettere di Anna, dopo la fine osservazione al microscopio fatta da Tullio De Mauro nel saggio d'introduzione alla prima edizione del volume, non hanno infatti ricevuto ulteriori approfondimenti linguistici, con l'eccezione di Sanga (2011) che ripercorre i tratti ascrivibili alle dimensioni di semplificazione, regolarizzazione analogica, ridondanza ed enfasi e provvede a rintracciarli nell'epistolario. Pare invece del tutto assente uno sguardo alla dimensione testuale e del discorso. A partire da queste considerazioni, attraverso un'indagine quantitativa e qualitativa, analizzeremo l'evoluzione del lessico nel corso della relazione epistolare fra Anna e Annabella Rossi, per poi focalizzarci sulla descrizione di fenomeni tipici della dimensione discorsiva (disfluenze testuali, uso di costruzioni a lista di tipo formulativo e denotativo, pragmaticalizzazione di una formula allocutiva, uso di strategie linguistiche tipiche del linguaggio familiare), frequentemente usati da Anna nelle sue lettere, in cui il dominio dell'oralità e quello della scrittura sono fortemente interconnessi.

2. La scrittura del popolo italiano e la tradizione dell'epistolografia

In ambito italiano l'utilizzo del materiale epistolare come oggetto di studio ha coinvolto, nel corso degli anni, diverse discipline e tradizioni. Accanto a studi di taglio più schiettamente letterario, focalizzati su epistolari spesso molto noti e

dedicati a scandagliare il rapporto tra scrittura privata, norma linguistica e prosa letteraria (cf. su tutti Mengaldo 1987; Guidolin 2011; D’Onghia 2014), buona parte dello studio sull’epistolografia condotta in territorio nazionale è debitrice del lavoro pionieristico di Spitzer dedicato alla corrispondenza dei prigionieri italiani della Grande Guerra detenuti nei campi di prigionia austro-ungarici (Spitzer 1976). Grazie al lavoro di Spitzer, introdotto in Italia nel 1976 con la traduzione di Renato Solmi, la linguistica decide così di dedicarsi all’epistolografia popolare, una “specie di letteratura popolare permanente che non prendiamo in considerazione solo perché ci è troppo vicina”, ma che permette di rispondere alla domanda “come scrive e cosa scrive il popolo” (Spitzer 1976: 1). L’attenzione linguistica a una forma di scrittura popolare ha quindi portato la comunità scientifica a concentrarsi sui testi dei cosiddetti scriventi semicolti, persone con scarsa o nulla scolarizzazione che per necessità hanno dovuto confrontarsi con la pagina scritta. Le contingenze storiche hanno sicuramente contribuito a rendere possibile questo filone di analisi: avvenimenti peculiari come i movimenti migratori nazionali e internazionali e le esperienze belliche hanno infatti prodotto materialmente una messe di testi (lettere, biglietti, autobiografie) prodotti da semicolti sui quali era possibile condurre l’analisi linguistica.

I corpora di corrispondenza diventano così un nuovo terreno di analisi dove confrontarsi per far emergere quell’italiano popolare teorizzato per la prima volta da De Mauro (2005 [1970]’) come il “modo di esprimersi d’un incolto che, sotto la spinta di comunicare e senza addestramento, maneggia quella che, ottimisticamente, si chiama lingua ‘nazionale’”. La grande quantità di studi condotta su testi di questo tipo rende sicuramente difficile una disamina esauriente sugli epistolari analizzati linguisticamente (si vedano ad esempio, senza pretesa di esaustività Rovere 1977; Caffarena 2005; Vanelli 2008). Giova però notare come lo studio degli epistolari prodotti da scriventi scarsamente alfabetizzati sia stato caratterizzato da un cambio di prospettiva. Gli studi condotti a partire dagli anni ’70 avevano infatti lo scopo prettamente diagnostico di osservare e riportare la sistematicità dei tratti di lingua degli scriventi, non ascrivibili in maniera esclusiva a nessun asse di variazione bensì frutto dell’intreccio tra scritto e parlato, dialetto e italiano standard, modelli colti e modelli popolari. Oggi, piuttosto che limitarsi a una esclusiva tassonomia volta a cogliere la devianza dalla norma o il dialettismo, si sugger-

isce uno sguardo che possa invece rendere conto del dinamismo di tali testi, collocandoli così “nel *continuum* di competenze scritte che sempre più chiaramente si va delineando allo sguardo degli storici della lingua, e che si va progressivamente sostituendo alla tradizionale opposizione italiano standard (letterario)/italiano popolare” (Fresu 2015: 15). I nuovi studi dedicati ai testi dei semicolti devono così tenere in considerazione la produzione scritta nel suo complesso, facendo emergere ad esempio le strategie testuali, le sfumature intermedie, il rapporto tra modelli testuali e linguistici (Fresu 2020).

In concomitanza con una maggiore e complessa attenzione alla produzione scritta semicolta risulta inoltre necessario focalizzarsi non tanto e non solo sul profilo dello scrivente, quanto sul genere testuale con cui lo scrivente si confronta, in modo da far emergere i suoi rapporti con eventuali modelli di riferimento nella costruzione dei cosiddetti “nuovi standard non standard” (cf. Allen 2015). La disamina alla *facies* linguistica del testo permette così di osservare non tanto e non esclusivamente le linee di tendenza dell’italiano che si evolve, quanto gli strumenti messi in campo per la negoziazione dei rapporti. Essendo la lettera uno dei luoghi privilegiati per la costruzione di un rapporto in assenza e per la sua negoziazione, è di cruciale importanza osservare le dinamiche linguistiche che permettono la costruzione di questo rapporto.

2.1 Lettere e costruzione del sé

Muovendo da queste considerazioni e accogliendo i suggerimenti di diversi settori della linguistica, in accordo coi loro diversi strumenti di analisi sia quantitativi sia qualitativi, è possibile un ripensamento nell’analisi di epistolari già analizzati per poter muovere considerazioni di aspetti finora lasciati da parte. In particolare, gli studi appartenenti al campo dei “New Literacy Studies” hanno più volte messo in evidenza come sia necessario osservare le pratiche di alfabetizzazione nel loro divenire e non solo nella loro staticità di processi mentali (cf. Gee 2015). Le lettere, assieme a una vasta e multiforme categoria di testi, sono pertanto un osservatorio per comprendere la conquista dell’alfabetizzazione intesa come pratica che riguarda non tanto i soggetti esclusivamente scriventi, quanto i parlanti intesi come soggetti sociali *tout court*. Scrivere una lettera deve essere quindi interpretato come una pratica socioculturale che sottintende altri fenomeni e altre pratiche socioculturali; secondo questa pros-

pettiva la lingua scritta non è mai del tutto slegata dalla lingua orale, bensì viene integrata in una pratica culturalmente situata, ed è da intendersi come uno dei modi possibili in cui la lingua svolge il suo ruolo pragmatico che permette di agire e di interagire (cf. Duranti 2005).

Gli studi condotti all'interno della cornice teorica dei “New Literacy Studies” hanno così messo in luce come, ad esempio, le lettere siano un modo per costruire e rappresentare l'identità di genere degli\ delle scriventi all'interno delle lettere d'amore, e come l'analisi delle lettere d'amore permetta di comprendere modelli culturali di produzione e riproduzione di determinate ideologie di genere (Baig & Naveed 2019). È stato inoltre notato come nel contesto migratorio, in cui vi è una distanza fisica spesso non colmabile, la lettera è uno strumento fondamentale per mantenere legami affettivi e per negoziare l'assenza dovuta alla migrazione (Gerber 2005, 2006; Borges 2016). In questa particolare situazione, ma non solo, la lettera diventa sia il simbolo materiale dell'assenza dell'altra persona sia la rappresentazione della sottoscrizione di un patto epistolare che, per essere valido, deve vedere la partecipazione in egual misura di entrambi gli\ le scriventi (Cancian 2011, 2012). La pratica scrittoria, vista come pratica culturale situata, diventa così un osservatorio che permette di comprendere in che modo le relazioni vengono costruite e rinegoziate, e come queste relazioni riflettono relazioni di potere presenti nella società. Per questa ragione, interpretando cioè, in linea con i “New Literacy Studies”, la lettera come una delle infinite forme possibili di materializzazione del valore pragmatico discorsivo del linguaggio, si può osservare il discorso presente nelle lettere inteso a tutti gli effetti come un'azione sociale mediata, in cui gli scriventi attingono a risorse linguistiche e a rappresentazioni socio-cognitive (come quella del sé e dell'altro) per stabilire, mantenere o sfidare relazioni di potere, muovere richieste, soddisfare compiti sociali e affettivi (Fairclough & Wodak 1997). Un'analisi linguistica condotta secondo questa prospettiva critica permette di far emergere, all'interno dei testi, relazioni di dominanza, di discriminazione, di potere, di controllo; può, inoltre, costituire un importante tassello per generare consapevolezza riguardo relazioni sociali di sfruttamento e subordinazione (Fairclough 2001; Burns & Carson 2005). A tale proposito la lettera risulta un genere testuale peculiare per un'analisi di questo tipo. Essa è infatti il luogo in cui la\ lo scrivente ha a disposizione la

sola pagina scritta per presentare il sé, per manifestare i suoi sentimenti e per evocare emozioni nei propri lettori (cf. tra gli altri Bednarek 2008).

2.2 *(Tra)scrivere l'oralità: fenomeni discorsivi nelle Lettere di Anna*

L'epistolario di Anna del Salento costituisce, un testo di sicuro interesse perché si presenta quasi come la trascrizione di un dialogo che la donna intrattiene con la destinataria delle sue lettere, presentando diverse caratteristiche sintattiche, semantiche, testuali e pragmatiche tipiche del parlato spontaneo. Se già la tradizionale opposizione tra scritto e parlato viene oggi considerata una dicotomia superata da una visione della lingua che prevede una vasta area di *continuum* tra i poli del parlato spontaneo e dello scritto altamente formalizzato, l'analisi delle produzioni semicolte mostra come "all'interno della stessa produzione scritta una gradualità verso il parlato può essere data dal parametro socioculturale, relativo al grado di istruzione dello scrivente" (D'Achille 2022: 83). Recentemente, Calaresu (2022) ha messo bene in luce come i testi prodotti in modalità scritta in diverse epoche e con diverse funzioni presentino ineludibilmente aspetti enunciativi e pragmatici oltre che fenomeni di tipo discorsivo, dal momento che l'interazione scritta consente e in un certo senso presuppone "una gamma ricorrente di attività metadiscorsive e di atteggiamenti nei confronti dell'interlocutore / destinatario", come ad esempio aprire o chiudere un discorso, introdurre un argomento o cambiarlo, ravvivare l'interesse del lettore, richiamarne l'attenzione, assicurarsene la benevolenza, e altro (cf. Calaresu 2022: 7). Nelle lettere scritte da Anna, in particolare, è possibile individuare diversi fenomeni linguistici e testuali che possono sembrare devianti, se paragonati alla norma di testi con modalità di produzione e ricezione tipiche dello scritto, ma che rappresentano invece veri e propri correlati funzionali della modalità parlata, dotata di caratteristiche altamente regolari e interlinguisticamente ben attestate.

3. Il corpus

Il corpus analizzato è costituito dalle lettere che Michela Margiotta, donna affetta da tarantismo e male di San Donato², ha inviato fra il 1959 e il 1965 all'an-

² Col nome di *male di san Donato* si identificavano in Salento sia le crisi epilettiche che pseudo-crisi di natura psicogena.

tropologa Annabella Rossi, raccolte e pubblicate da quest’ultima per la prima volta nel 1970 nel volume *Lettere da una tarantata*, col saggio introduttivo di Tullio De Mauro del quale si è scritto. Il volume è stato riedito da Paolo Apolito per Squilibri: a quest’ultima edizione si fa riferimento nel corso di questo lavoro, anche nei rimandi al testo delle lettere. La scrivente è stata nel volume pubblicato anonimizzata con lo pseudonimo di Anna, adottato anche in questo lavoro (perché è quello effettivamente presente nel testo in analisi). Per alcuni dettagli sulla nascita dello scambio fra le due donne e una descrizione tematica dell’epistolario rimandiamo a Nodari & Corona (in stampa).

L’epistolario contiene 65 lettere, di cui 44 scritte personalmente da Anna e 21 dettate a scrivente più esperta (secondo una pratica diffusa e ben documentata tra le persone non scolarizzate); in Tabella 1 si riportano alcune informazioni relative al corpus.

	Tokens complessivi	Lunghezza media lettere
Lettere autografe	25150	571.6 (s.d. 390.7)
Lettere dettate	5962	283.9 (s.d. 159.7)

Tabella 1. Descrizione quantitativa del corpus.

Nelle edizioni dell’epistolario le lettere sono riportate in ordine cronologico e l’indicazione “Lettera dettata” accompagna in nota tutte le lettere non scritte da Anna. Una prima osservazione stilistica delle lettere dettate porterebbe a ipotizzare la presenza di due diverse scrivane, con diversi livelli di competenza scrittoria; tuttavia, quest’osservazione andrebbe corroborata con analisi stilometriche che non condurremo in questa sede. Inoltre, l’assenza delle lettere autografe originali (di cui non è possibile ricostruire la storia di conservazione) non ci permette di verificare questa ipotesi sulla base delle grafie presenti.

L’analisi condotta in questo lavoro si concentrerà, in particolare, sul subcorpus costituito dalle lettere scritte da Anna; tuttavia, come spiegheremo nel paragrafo seguente, abbiamo condotto specifici rilievi, dove necessario, anche sulle lettere dettate.

4. Un’analisi quantitativa delle lettere di Anna

Per l’analisi linguistica abbiamo innanzitutto costruito due subcorpora: il primo corpus, composto da 44 lettere, corrisponde alle lettere scritte inter-

amente da Anna, mentre il secondo corpus, composto da 21 lettere, corrisponde invece alle lettere che Anna ha delegato ad altre scrivane di Ruffano. L'analisi lessicale è stata quindi condotta di volta in volta separatamente, in modo da far emergere la singolare pratica scrittoria di Anna.

Una parte dell'analisi è stata condotta attraverso il software Voyant, una piattaforma d'accesso gratuita e *open source* messa a punto per l'analisi testuale e per il *text mining* (<http://voyant-tools.org>). Grazie all'interfaccia di Voyant è possibile osservare i caratteri generali dei documenti studiati e notare alcune linee di tendenza su cui possono poi essere condotte delle analisi *ad hoc*.

Un primo sguardo a dei dati quantitativi ci può innanzitutto dare delle informazioni rispetto alla ricchezza di vocabolario di Anna. Si osserva così che nell'intero corpus di lettere autografe il rapporto *type/token* (TTR) è molto basso. Per quanto il TTR sia una misura globale e pertanto solo indicativa, si osserva che all'interno delle 44 lettere compaiono infatti 25150 *tokens* ma solo 3243 *types*, corrispondenti a un TTR del 12%. Il valore si apprezza di più se si considera il TTR per le lettere scritte da mano più esperta, che ha un TTR del 27% e conferma la maggiore ricchezza di vocabolario. Non dissimilmente, il calcolo dell'indice *Gulpease*, basato sulla lunghezza delle parole (in termini di numero di lettere), sul numero delle parole e sulla lunghezza delle frasi, conferma la bassa leggibilità delle lettere, dal momento che le lettere autografe di Anna hanno un indice medio pari a 52³; notabilmente, il confronto con il subcorpus delle lettere dettate conferma ancora una leggibilità maggiore di queste ultime, le quali hanno un indice medio pari a 66.

Voyant offre inoltre la possibilità di considerare la densità del vocabolario sia dell'intero corpus, sia dei singoli documenti: un alto valore di densità di vocabolario indica che all'interno di un testo ci sarà un'ampia variazione lessicale. A questo proposito si nota come la densità del vocabolario delle lettere di Anna pare avere un andamento crescente. La prima lettera risulta

³ Per una descrizione dettagliata dell'indice *Gulpease*, cf. Lucisano & Piemontese (1988). L'indice ha dei valori compresi tra 0 e 100, con il valore 100 corrispondente a una leggibilità più alta e 0 a una leggibilità più bassa.

essere infatti in assoluto la più lunga, ma anche quella con minore densità di vocabolario (1936 *tokens*, 604 *types*, con un TTR pari al 31%) mentre, al contrario, la lettera più densa di vocabolario pare essere l’ultima lettera (con una densità pari al 61%). Certo è che la scarsa dimestichezza di Anna con la pratica scrittoria e la sua incompleta scolarizzazione si notano anche nel momento in cui il corpus delle lettere autografe viene confrontato con le 21 lettere dettate a scriventi più esperte. In questo caso la massima densità di vocabolario viene raggiunta nella lettera 57, con un TTR dell’83%. Purtuttavia, come osservato prima, seppur le lettere dettate abbiano un valore medio di TTR più alto rispetto alle lettere autografe (27% vs 12%), la densità non è comunque particolarmente elevata dato il tono convenzionale e formulaico che si osserva nelle lettere dettate.

Attraverso lo strumento Cirrus implementato in Voyant è possibile avere una prima idea delle parole più frequenti nell’intero corpus. In una prima fase esplorativa si è osservato il corpus nella sua interezza, senza applicare il filtro delle *stop words* (cf. Vitali 2020 per una strategia analoga sulla scrittura dei semicolti), come si nota nella Figura 1a. Successivamente la nuvola è stata generata dopo aver selezionato una lista di *stop words* già disponibile per l’italiano all’interno del programma stesso, che permette così di filtrare parole funzionali come articoli, congiunzioni e preposizioni (Figura 1b). Infine è stato scelto di applicare la lista di *stop words* già disponibile in Voyant, a cui sono stati però sottratti i pronomi personali e gli aggettivi come *mia* e *vostra*, poiché frequenti nelle forme di saluto e significativi della scrittura allocutiva di Anna (Figura 1c, v. più avanti). In Tabella 2 vengono riportate le 5 parole più frequenti per ogni condizione poc’anzi dettagliata.



Figura 1a. Parole più frequenti nel corpus delle lettere scritte da Anna, senza filtrare le *stop words*.



Figura 1b. Parole più frequenti nel corpus delle lettere scritte da Anna, filtrando le *stop words* tramite la lista presente in Voyant.



Figura 1c. Parole più frequenti nel corpus delle lettere scritte da Anna, eliminando dalla lista delle *stop words* alcuni pronomi personali e aggettivi caratteristici della scrittura allocutiva di Anna.

	Prime cinque parole più frequenti
Senza <i>stop words</i>	che (1148); e (1071); di (782); la (771); a (636)
Con filtro delle <i>stop words</i>	signorina (277); buona (247); Anna (241); cara (159); quando (153)
Con filtro <i>ad hoc</i>	mia (495); io (390); signorina (277); buona (247); Anna (241)

Tabella 2. Parole più frequenti con o senza *stop words*.

Applicando la lista di *stop words* così modificata, le parole con la più alta frequenza numerica nell'intero corpus di lettere autografe risultano essere *mia*, *io*, *signorina*, *buona*, *anna*, *cara*. In virtù delle peculiarità della scrittura di Anna, caratterizzata da segmentazioni scorrette o univerbazioni, nella lista delle parole meno frequenti compaiono anche esempi di discrezione dell'articolo (es. *zione*, attestato una sola volta nella frase *elle ragazze della zione cattolica anno fatta una*

bellissima recita dentro aunolocale, [L2, p. 115]), di errata segmentazione (es. una singola occorrenza di *zidera*, in *con zidera la Vostra Cara Anna quantto a soferto di tutte le mali sofferenze*, [L3, p. 120]) o di concrezione del pronome clitico (es. una singola occorrenza di *tiracomando*, in *e quello che tiracomando di scrivere presto*, [L38, p. 160]).

In Tabella 3 vengono riportati, oltre alla frequenza assoluta, anche i valori di acutezza (*peakedness*) e di *skew* delle parole più frequenti nell'intero corpus. Da questo punto di vista le misure di acutezza e di *skew* possono dare delle indicazioni relative all'andamento della corrispondenza epistolare. L'acutezza è infatti una misura statistica che indica quanto le frequenze relative di un termine all'interno di un corpus sono addensate in picchi. Un alto indice di acutezza indica che vi sono regioni del corpus in cui una parola ha valori di frequenza più alti rispetto ad altre sezioni del corpus: essendo il corpus organizzato in ordine cronologico, un cambio nei valori di acutezza può significare un'evoluzione nello scambio epistolare, con cambiamenti nelle scelte lessicali. Non dissimilmente dall'acutezza, la *skew* mostra la simmetria delle frequenze relative di un termine. Ciò significa che un termine che compare in maniera omogenea in tutti i documenti di cui è composto il corpus avrà un indice di *skew* vicino allo zero. In questo caso si nota che un termine come *cara* pare essere meno simmetrico rispetto ad altre parole che sono contraddistinte anche da un basso valore di acutezza come *mia*, presente omogeneamente nell'intero corpus di lettere manoscritte. Ovviamente i valori più alti di acutezza si ritrovano sia per gli hapax, come per la parola *papa*, che compare solo nella lettera 52 e ha un'acutezza di 44, sia per parole altamente frequenti ma che tendono a comparire in specifiche regioni del corpus. Si nota così come due parole molto frequenti come *buona* e *vostra* paiono addensarsi soprattutto nelle prime lettere, tanto che nella sola lettera 5 compaiono ben 6 occorrenze di *buona* e 7 occorrenze di *vostra*, come riportato di seguito:

(1) *Mia Buona Signorina quando vieni di S'antopaolo per favore mi porti il barattolo [...] tiò preparata una cosa buona che vuoi mangiate colla vostra Madre è pure colla vostra sorella è pure il vostro Papà se la vostra Anna potrà avere onori di tutta la vostra famiglia [...] la vostra buona Signorina mia la vostra foto lo messa a un bellissimo quadretto vicino a me [...] ti guardo sempre perche siei buona come mi mandi la foto della mia buona Mamma che la voglio vedere che a me è una Buona Madre saluti alla mia Mamma della sua figlia Anna ciao ciao.* [L5, p. 122]

Parola	Conteggio	Acutezza	Skew
mia	495	0.1	0.2
io	390	-1.1	-0.1
signorina	277	0.7	0.9
buona	247	2.1	1.1
anna	241	-0.8	0.4
vostra	188	2.4	1.5
cara	159	3.4	1.6
quando	153	2.8	1.1
casa	123	3.4	1.3
voi	104	2.0	1.3
me	99	2.8	1.1

Tabella 3. Valori di acutezza e di skew delle parole più frequenti nell'intero corpus.

L'osservazione relativa all'acutezza e alla skew viene confermata nel momento in cui si considerano le linee di tendenza, un modo efficace per rendere graficamente alti valori di acutezza e di skew. Nella Figura 2 si osserva l'andamento nel corpus della parola *cara*, parola altamente frequente che pare però essere concentrata nella parte finale dell'epistolario. Significativamente, questo termine affettivo ma più convenzionale mostra la massima frequenza proprio nell'ultima lettera dell'epistolario, quella con la quale si chiude in maniera dolorosa la corrispondenza epistolare con Annabella Rossi (es. 2).

(2) Mia Cara Signorina non puoi mai immaginare cheò ricevuta il vostro caro scritto che la tua cara Anna che sono stata tanto con pensiero da quando tene-si andata della mia casa [L64, p. 183]

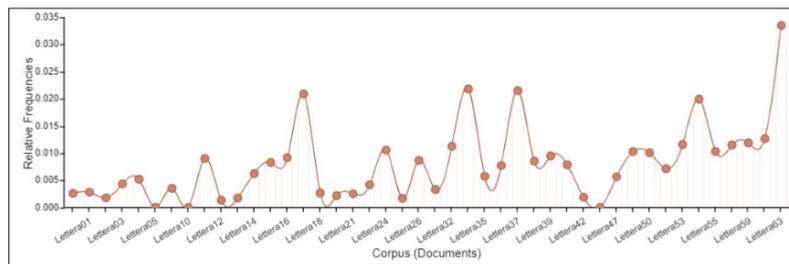


Figura 2. Andamento nel corpus della parola *cara*.

L’andamento frequenziale crescente di *cara* pare inoltre avvenire a discapito di *buona*, termine che Anna spesso usa in alternanza nelle sue tipiche forme di apertura e saluto epistolare. Nella Figura 3 si nota che nelle lettere pare avvenire una graduale sostituzione: se nelle lettere iniziali la parola *buona* sembra essere la più frequente, nel proseguimento della corrispondenza epistolare questo aggettivo pare ritirarsi in favore della forma più standard *cara*.

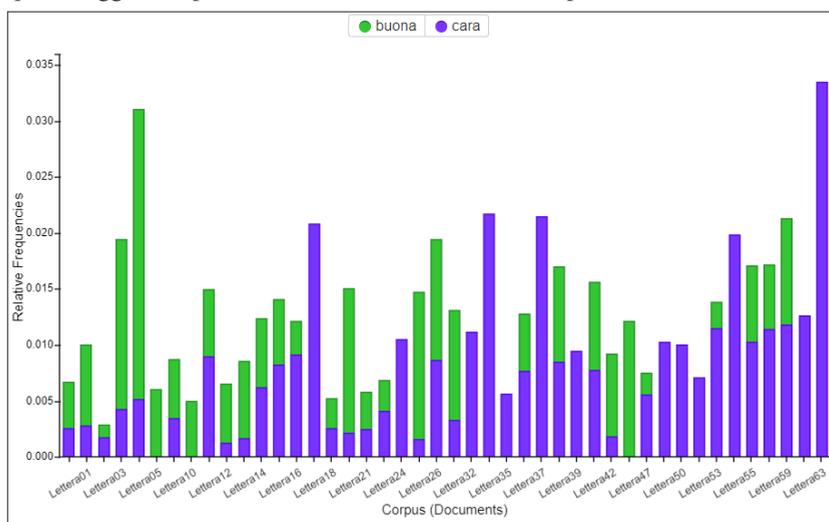


Figura 3. Andamento nel corpus della parola *cara* in rapporto a *buona*.

I coefficienti di correlazione, calcolati secondo il test di correlazione di Pearson, confermano che comunque, per buona parte dell’epistolario, la formula preferita da Anna risulta essere *mia buona*, con i due termini moderatamente correlati ($r=0.5, p<0.05$)⁴. *Cara* pare comunque altrettanto presente in congiunzione con

⁴ Il calcolo del coefficiente di correlazione di Pearson viene effettuato da Voyant grazie alla presenza di una libreria Apache Math Commons che contiene la funzione SimpleRegression. Il coefficiente di correlazione viene calcolato confrontando le frequenze relative dei termini. Un valore prossimo a 1 indica una correlazione positiva, con i due termini che tendono a comparire assieme, mentre un valore vicino a -1 indica una correlazione inversa, con la presenza di un termine a discapito dell’altro della coppia; infine, valori prossimi allo zero rappresentano una scarsa correlazione.

mia, dato che anche i due termini paiono correlare, seppur lievemente ($r=0.3$, $p<0.05$). Una ricerca dei sintagmi che più frequentemente compaiono assieme conferma comunque l'analisi: il sintagma *mia buona signorina* risulta essere il più frequente di tutto l'intero corpus di lettere autografe (65 occorrenze), mentre *mia cara signorina* mostra 19 occorrenze. Uno sguardo alle correlazioni con *signorina* conferma ancora una volta una maggiore correlazione con *buona* ($r=0.4$, $p<0.00$) piuttosto che con *cara*, la cui correlazione è debole e non significativa ($r=0.2$).

Significativamente, l'osservazione sull'alternanza di *buona* e *cara* nel lessico di Anna viene corroborata dall'analisi descrittiva condotta sul subcorpus delle lettere dettate. Nelle 21 lettere è infatti *cara* a essere prevalente su *buona* a livello frequenziale. *Cara* è infatti il terzo termine più frequente dell'intero subcorpus, con 55 occorrenze, mentre *buona* compare solo al dodicesimo posto in ordine decrescente, con 19 occorrenze. Nel corpus delle lettere dettate *cara signorina* è inoltre il sintagma più frequente (21 occorrenze), mentre *mia buona signorina* mostra 8 occorrenze. Nel caso delle lettere dettate la correlazione più forte si osserva tra i termini *cara* e *signorina* ($r=0.6$, $p<0.00$), mentre i termini *buona* e *signorina* mostrano una correlazione non significativa.

In generale, l'analisi delle frequenze del subcorpus di lettere dettate conferma la presenza di un linguaggio fortemente formulaico e convenzionale. Come si nota in Tabella 4, in questo caso le parole più frequenti risultano essere *signorina*, seguita da *cara*, *mia*, *io*, *Anna*. Ancora una volta quindi le parole più frequenti risultano essere quelle usate in apertura e chiusura di lettera.

Parola	Conteggio	Acutezza	Skew
signorina	86	-0.6	0.0
cara	55	0.0	0.7
Mia	54	1.1	1.0
Io	34	1.8	1.3
anna	30	0.0	0.5
tanto	28	0.1	1.0
mio	27	-0.8	0.6

Tabella 4. Valori di acutezza e di skew delle parole più frequenti nel corpus delle lettere dettate.

4.1 Il lessico di Anna e le parole per parlare di sé

Dopo un primo sguardo d'insieme del lessico di Anna è possibile osservare più attentamente alcuni fenomeni specifici. Come già detto in relazione al

rapporto tra *types e tokens*, il lessico delle lettere autografe pare essere ridotto e poco variato, tanto che il nucleo del lessico di Anna pare essere caratterizzato dalle formule di apertura e chiusura dell’epistolario. È interessante osservare però, come già riportato in Tabella 1, l’alternanza tra i termini che Anna sceglie per parlare di sé e per mettere al centro dello scambio epistolare l’assenza di Annabella e la negoziazione del rapporto. Il pronome personale *io* è infatti il secondo termine più frequente (390 occorrenze), ma un termine molto frequente pare essere anche *Anna*, il suo stesso nome proprio. I due termini paiono inoltre correlare debolmente ($r=0.2$). In termini assoluti, è nella prima lettera che il pronome personale *io* compare con alta frequenza (40 occorrenze). In questa lettera Anna è infatti chiamata a presentarsi e a raccontare le proprie esperienze di vita (es. 3):

(3) che *io* sono nata sportonata è devo morire sí. [...] mio padre aveva una grossa proprietà e *io* era più piccola di tutte e *io* andava spesso alla chiesa e *io* siccome che non voleva stato di matrimonio li diceva al mio Padre di comprare una casetta ma mio Padre faceva lorecchio di marcantte e *io* lo sucerviva alla Mamma [...] mio Padre era fatto il testamento *io* hò penzato che mia lasciata la casa a me e *io* li diceva alla mia Madre di non dare tutte cuelle bestemie [L1, p. 109]

Nell’arco dell’epistolario, il pronome personale viene usato per riportare i momenti significativi della vita di Anna, come il suo primo incontro con il cinema, o per raccontare le sue azioni quotidiane (ess. 4-5), ma anche per riportare stralci di discorso diretto, in una sorta di mimesi del parlato (es. 6):

(4) *io* quando andai la prima volta allo Cinoma vetti cose che non lo creteva come è ma quando lo vetti coi miei occhi *io* vetti cose mai viste alla mia vita [L2, p. 114]

(5) *io* alla mattina mi alzava un po presto e metteva la pignata dei piselli [L3, p. 118]

(6) *io* lo risposto è lo detto che non si netrovano come li voi tù S’andonato mia risposto e adetto vai alla coprativa che li trovi come li voglio *io* e *io* lo risposto a S’andonato [L2, p. 113]

È però nel segnalare il rapporto affettivo con Annabella e anche la disparità di coinvolgimento nel rapporto che il pronome personale *io* compare

frequentemente nelle lettere. In questi casi Anna mette spesso sé al centro, ribadendo l'importanza dello scambio epistolare con l'antropologa (ess. 7-8):

(7) la tua Anna che ti vole tantto bene che voi pure li voi tanto bene che *io* non tengo nessuno altro che voi [L1, p. 112]

(8) che *io* a spetava quella cosa che voi eri detto per una bella sopresa ma la tua Cara Anna non ha ricevuto propio niente [L59, p. 177]

Proprio quest'asimmetria di relazione si osserva nei numerosi casi in cui Anna si appella col proprio nome per parlare di sé e presentarsi come soggetto investito di esperienze dolorose, soprattutto in rapporto alla relazione con Annabella. Come già osservato in Tabella 1 *Anna* è anzi una delle parole più frequenti nell'intero corpus, e si ritrova in specifici contesti di occorrenza, spesso tra loro in sovrapposizione. Il termine è innanzitutto presente nelle formule di saluto in chiusura, in cui Anna si presenta, non solo firmando con il suo nome, ma anche definendosi come soggetto amoroso in relazione con Annabella (ess. 9-10):

(9) E Sono per Sempre Vostra Cara *Anna* [L3, p. 120]

(10) ricevi della vostra Cara *Anna* a Voi Mia Buona Signorina mille e mille strette di Mano da lontano [L16, p. 135]

Seguono poi casi di riproduzione di discorsi, in cui il nome proprio compare nei discorsi citati (Calaressu 2004, es. 11):

(11) mianno risposto *Anna* a qui ce un giovane alto e sottile [L24, p.]

I casi più numerosi e più peculiari sono però quelli in cui Anna fa uso di una strategia riconducibile al cosiddetto "illeismo", inteso come una strategia retorica che consiste nel parlare di sé in terza persona. In questi contesti il riferirsi al sé viene compiuto da Anna con il referente personale per antonomasia, ossia il nome proprio, spesso accompagnato da descrittori (ess. 12-13).

(12) ma la tua *Anna* non a cettava propio per niente che mera detto il demonio [L1, p. 111]

(13) vidi che la tua Cara *Anna* mancherà nove giorni della mia casa [L10, p. 126]

Vi sono infine casi in cui la scrivente si identifica chiaramente come destinataria dell'azione. In questi casi l'utilizzo del termine *Anna* può comparire sia all'interno della semantica e della sintassi della frase (es. 14), sia in maniera staccata (es. 15), echeggiando così casi di vocativo inverso tipici dei dialetti centromeridionali (Rohlf's 1966, ma vedi par. 5.4).

(14) non fare rabiare la Vostra *Anna* che mi sento tanto male [L16, p. 134]

(15) dimi come sono state le mie lettere che teo mandate la tua *Anna* [L11, p. 128]

Si può quindi notare come, almeno a livello quantitativo, non sia facile riscontrare un vero e proprio lessico affettivo nelle *Lettere*. Come detto, i primi sette termini più frequenti possono infatti essere ricondotti alle formule di saluto di apertura e chiusura che *Anna* usa frequentemente (*mia, signorina, buona, Anna, vostra, cara*) o al pronome personale usato per riportare i discorsi. Purtroppo, soprattutto nei casi in cui *Anna* entra a pieno titolo come protagonista dell'enunciazione menzionando sé stessa come destinataria dell'azione, si osserva il coinvolgimento di *Anna* nel processo scrittoriale e di mantenimento del rapporto.

Al di là dei riferimenti a sé, uno dei primi termini pieni che si ritrova nel lessico di *Anna* è *casa*, in relazione alle numerose volte in cui racconta il rapporto travagliato con la sua famiglia, ma significativamente compare in maniera abbastanza frequente anche il termine *male*. La parola ribadisce l'originaria natura dello scambio epistolare, nato proprio in seguito all'interesse di *Annabella Rossi* per i fenomeni di tarantismo e del cosiddetto *male di san Donato*. Da questo punto di vista, la parola *male* compare infatti frequentemente all'interno del sintagma *male di s'andonato*. Negli esempi 16-17 si osserva come *Anna* utilizzi la parola sia per riferirsi specificamente al *male di san Donato*, sia più genericamente agli attacchi epilettici riconducibili al tarantismo.

(16) voi sapere come hò capito il *male* di S'andonato [L2, p. 113]

(17) che io ne soffro il *male* piletica e mi disse cosa e questo *male* io li dissi il *male* di San Donato e lei disse povera signorina e mi disse vieni se tu avrai questo *male* se ti viene ti aiutiamo noi [L23, p. 143]

Sebbene il lemma compaia con più frequenza in quelle lettere il cui argomento è relativo al tarantismo e al *male di san Donato*, il tema del *male* è però

presente in maniera latente in tutta la corrispondenza. Le sofferenze affrontate da Anna non sono solo in relazione al male di san Donato, ma rimandano più genericamente alla sua vita di stenti e difficoltà e alle sue sofferenze fisiche (es. 18):

(18) non pasai più niente di quello *male* questa la mia vita che hò pasata tanto *male* che non o cotuto mai bene da lora che sono nata mo per uno e mo per laltro guai sono stata sempre *male* [L1, p. 112]

Per Anna c'è però una ulteriore causa di sofferenza che si sposa ai dolori già patiti per gli attacchi epilettici dati dal male di san Donato, ed è la situazione conflittuale che ha con Annabella. Osservando i contesti di occorrenza del termine si nota infatti come Anna spesso utilizzi la parola per riferirsi all'asimmetrico rapporto instauratosi con l'antropologa, che è per lei fonte di rabbia e frustrazione (ess. 19-20):

(19) sono rimasta tanto *male* a sentire che sarà difficile di venire di S'ampaolo non puoi credere come mi sono sentita *male* perché io vedendo che tardi a crivere immaginavo che mi facevi una sorpresa [L42, p. 164]

(20) mi fai venire i nervi di più che li tengo tanti brutti e sono di piu dempestosa che mi fai sentire *male* la Vostra Cara Anna [L62, p. 181]

In conclusione, l'analisi lessicale mostra come, seppure in possesso di un vocabolario non particolarmente esteso, Anna sia comunque in grado di utilizzare le abilità linguistiche per negoziare di volta in volta la relazione con Annabella: i suoi limitati strumenti le permettono di porsi al centro di un nodo narrativo in cui Anna può continuare a tessere il suo rapporto spesso conflittuale con Annabella.

5. La dimensione discorsiva nelle *Lettere di Anna*

I rilievi fatti sul lessico delle lettere di Anna possono essere meglio osservati tenendo conto di un assunto di base, anticipato già in 2.2: chi non ha avuto uno specifico addestramento alla scrittura, pur essendo in qualche modo alfabetizzato, resta sempre fortemente ancorato alla sfera dell'oralità (cf. fra gli altri D'Achille 2022: 83). Il legame con la dimensione discorsiva viene spesso usato per caratterizzare la sintassi dei testi prodotti da scriventi scarsamente scolar-

izzati, in cui si rileva un andamento “dominato da paratassi e giustapposizioni, che denuncia modalità argomentative molto più vicine all’oralità” (Volpi 2010: 126). In realtà, come le lettere di Anna testimoniano, sono numerosi i correlati funzionali della modalità parlata nei testi prodotti da scriventi semicolti, non solo sintattici ma anche semantico-funzionali. Ne presenteremo di seguito quattro che caratterizzano lo scritto di Anna: la presenza di fenomeni di disfluenza testuale, come la ridondanza tematica, e di “plurideterminazione” semantica (in 5.1); l’impiego di liste denotative con diverse funzioni (in 5.2); la pragmaticalizzazione di formule che vengono usate, nel testo, come segnali discorsivi con valore metatestuale o interazionale (in 5.3); l’uso di forme di vocativo e allocuzione inversa e di illeismo tipiche del linguaggio familiare (in 5.4).

5.1 “[I] punti e le virgole le metterai sola dove ci vogliono”: *paratassi, giustapposizione e ridondanza*

L’analisi condotta con Voyant ha già permesso di rilevare come tratti tipici della scrittura di Anna la ridondanza e le frequenti ripetizioni: queste caratteristiche sono inserite in una costruzione della frase che si sostanzia in una giustapposizione di sintagmi in sequenza. Analizziamo alcune caratteristiche sintattiche e semantiche tipiche della costruzione dell’enunciato nell’epistolario, osservando gli esempi da (21) a (24).

(21) questa la mia vita che hò pasata tanto male che non o cotuto mai bene da lora che sono nata mo per uno e mo per laltro guai sono stata sempre male ma non solo questo ma anche le maldicenze delle altre bisognato di somporare⁵ tutti li anni che sto sola e per questo la tua Anna che ti vole tantto bene che voi pure li voi tanto bene che io non tengo nessuno altro che voi [L1, p.112]

(22) non hò potuto Mandare più presto la mia lettera perché non sono stata bene che sono stata male a letto che la Maria mia fatta belinare tanto mia dette tante cattive parole e io mi sono stato in solenzio non ò risposto per niente non creti come sto male non mangio e non bevo e se me vedi come sono ridotta specialmente a desso che voi lo sai come sto male per la taranta che non mi fa mangiare niente che morio della fame [L2, p.115]

⁵ Nell’edizione consultata, la voce viene interpretata come possibile variante di *superare*, sebbene sia considerata un’interpretazione dubbia (accompagnata infatti dalla notazione [?]). Più probabilmente la voce va intesa come variante di *sopportare*.

(23) le mie vicine mianno detto che domandao di Maria se ancora sto come prima e io dissi che nera potuto fare meno di questa cosa che io non posso fare mai più pace di quello che a me mia fatto che voi Mia Buona Signorina non puoi sapere tutto di quella da quantte mena fatte a me di tutti colori a quello che io lo fatto a Maria del bene che io li hò fatto da tantti anni che io doveva parare da tutto quello che lui voleva della mia casa [L16, p.133]

(24) io hò cretuto che tu mi rispondi subito alla Mia lettera ma ho cretuto che tu verso della tua Anna mi fai sempre piangere che tigo come sono io per te non siei tu per me che sinò un rigo che datanto tempo che non ricevo una vostra posta vorrei⁶ se non ti batte il vostro cuore [L40, p.162]

Negli estratti dalle lettere qui proposti risulta evidente una delle caratteristiche spesso sottolineate in questi testi, cioè la totale assenza di segni di punteggiatura che rende la lettura scarsamente agevole. L'uso dell'interpunzione è in tutte le scritture semicolte "estremamente caotico" (D'Achille 2002: 115), in alcuni casi quasi del tutto assente, al punto che Cortelazzo (1972: 119) individuava in queste produzioni un effetto di "smarrimento interpuntorio".

Questo avviene non soltanto in virtù della scarsa scolarizzazione e dello scarso dominio della punteggiatura ma anche a causa della modalità di costruzione del testo che sembra programmata *online*, attraverso il tipico meccanismo di giustapposizione paratattica che rappresenta uno dei principali correlati sintattici dell'oralità. Inoltre, come si può osservare negli estratti in (21)-(24), anche la costruzione della referenza presenta tutte le caratteristiche di quello che Voghera (2017) definisce "un risultato semantico che ha come sua proprietà principale la bassa definizione". Anche da un punto di vista semantico, infatti, la forma sembra definirsi man mano, perché chi scrive non sembra "sempre coscientemente occupato nella scelta della parola o dell'espressione esatta o calzante, ma costruisce il significato strada facendo" (Voghera 2017: 163). In questo meccanismo di costruzione

⁶ Qui possiamo ipotizzare l'omissione non controllata del verbo *sapere* perché, nella stessa lettera, poco prima, Anna – il cui stile è, come mostrato, ripetitivo e spesso formulare – scrive la stessa frase in forma completa: «ti scrivo questi pochi rigi per far sapere cheme va questo fatto che non mi rispondi alla tua Anna che non so cosa devo dire più che non ricevo ne suno rigo di voi come va questo fato? che non mi fai prendere riposo ne sun momento della mia vita vorrei sapere se non ti batte il tuo cuore per la tua Anna».

della referenza, tipico del parlato interazionale, il significato va definendosi man mano perché i parlanti, nell’atto comunicativo, si affidano più o meno consapevolmente alla cooperazione col destinatario. I testi prodotti nel parlato dialogico sono infatti caratterizzati da temi sospesi, esitazioni (che hanno una controparte di significante in pause piene o fenomeni di disfluenza fonetica)⁷, false partenze o cambi di progetto. I testi prodotti in modalità parlata sono infatti “inerentemente” aperti alla possibilità di ridefinizione grazie all’intervento di un destinatario, che coopera alla costruzione della referenza apportando costantemente un proprio contributo conversazionale (cf. ancora Voghera 2017: 191 e ss.).

Per quanto testimoniato dalle lettere, nella scrittura di Anna la destinataria da un lato viene invocata come se fosse effettivamente partecipe dell’atto discorsivo, dall’altro viene chiamata a cooperare nella costruzione della referenza che si definisce *on line* perché le limitate risorse lessicali a disposizione della scrivente vengono interamente impiegate per raggiungere una descrizione della realtà che possa essere condivisa. Questo tipo di processo di costruzione del testo, da un punto di vista sintattico e semantico, è reso evidente in (21)-(24) dai fenomeni di disfluenza testuale tipici della modalità parlata. Le disfluenze testuali sono quelle che non alterano la catena fonica ma interrompono la sequenza verbale e richiedono “di norma, un’operazione di ricostruzione testuale da parte dell’ascoltatore” (Voghera 2017: 73). Nel caso dei testi prodotti da Anna, a essere chiamata a un’opera di decodifica, a fronte di una testualità fortemente sconnessa, è la destinataria delle lettere.

In (24), ad esempio, possiamo trovare in rapida sequenza due diversi fenomeni di disfluenza testuale rappresentati da cambi di progetto (“ho cretuto che tu verso della tua Anna | mi fai sempre piangere”; “come sono io per te non sei tu per me che sinò un rigo che datanto tempo che non ricevo una vostra posta | vorrei [+sapere?] se non ti batte il vostro cuore”).

Altra caratteristica molto evidente nelle lettere di Anna, e ben rappresentata in (21)-(24), è la ridondanza. La ridondanza, una delle caratteristiche più tipiche del parlato, ha infatti una precisa funzione nei processi di comunica-

⁷ Un’esaustiva rassegna dei principali fenomeni di disfluenza fonetica è in Heldner & Heldlund (2010).

zione orale: serve ad assicurare una funzione di aggiustamento e recupero di eventuali pezzi di informazione che, nella pratica discorsiva, possono andare persi a causa di condizioni connaturate al parlato dialogico (mancanza di programmazione da parte del parlante, necessità di richiamare l'attenzione del destinatario, calo di attenzione da parte del locutore, interruzioni dovute a varie forme di rumore ambientale, cf. su questo già Sornicola 1981). Nello scritto di Anna, questo correlato funzionale della modalità parlata è presente in maniera evidente, pur non essendo connaturato alla modalità di trasmissione effettivamente adottata.

Nelle lettere è possibile osservare, in particolare, il tipo di ridondanza tematica che porta alla produzione di strutture e strategie testuali molto varie: si va da semplici forme di ripetizione a riformulazioni, parafrasi e progressivi ampliamenti del tema attraverso forme di riscrittura di concetti già espressi⁸.

De Mauro (2005 [1970]¹: 80) sosteneva che la sintassi di Anna, “a prima vista deforme”, acquistava un andamento più leggibile “con semplici restauri della grafia e della punteggiatura”. Altra soluzione, che permette anche di apprezzare in tutta la loro evidenza i fenomeni sintattici e testuali appena descritti, è adottare il modello a griglia proposto da Blanche-Benveniste *et al.* (1979) per l'analisi di testi di francese parlato⁹. Il modello a griglia, usato fino ad oggi solo per testi di parlato spontaneo, offre una presentazione grafica del testo che permette di seguire più agevolmente la costruzione della referenza per come avviene nell'interazione linguistica. Mentre nella costruzione del testo scritto si sfrutta, infatti, soltanto la dimensione sintagmatica, che si sviluppa orizzontalmente, nell'enunciazione spontanea i parlanti esplorano cognitivamente anche la dimensione paradigmatica e i testi si costruiscono sfruttando anche un asse di sviluppo verticale, (cf. Blanche-Benveniste 1990).

Le lettere di Anna sono sintatticamente costruite proprio sulla multidimensionalità che caratterizza il parlato spontaneo: da un lato, il testo si svi-

⁸ Questi casi di riformulazione sono stati definiti, nel campo degli studi conversazionali, come esempi di completamento (*completion*) o di auto-correzione (*self-repair*), fenomeni ampiamente diffusi nel parlato spontaneo interazionale, cf. *inter alia* Lerner (1994).

⁹ Il modello di descrizione a griglia è stato usato anche nella descrizione di fenomeni del parlato italiano in alcuni lavori, fra i quali Bonvino *et al.* (2009), Masini & Pietrandrea (2010) o Voghera (2017).

luppa orizzontalmente, attraverso sequenze sintagmatiche di costituenti che rappresentano la narrazione sequenziale; verticalmente, invece, troviamo inseriti in colonna tutti i costituenti che realizzano fra loro rapporti paradigmatici e che rappresentano le esitazioni, le riformulazioni, le parafrasi e gli ampliamenti tematici tipici della scrittura di Anna, fortemente imperniata sull’oralità.

1.	questa la mia vita			
2.		che hò pasata tanto male		
3.		che non o cotuto mai bene da lora che sono nata		
4.		mo per uno e mo per laltro guai sono stata sempre male		
5.		ma non solo questo		
6.		ma anche le maldicenze delle altre bisognato di somporare tutti li anni che sto sola		
7.			e per questo la tua Anna	
8.				che ti vole tantto bene
9.				che voi pure li voi tanto bene
10.				che io non tengo nessuno altro che voi

1.	non sono stata bene			
2.	che sono stata male a letto			

3.		la Maria mia fatta belinare tanto			
4.		mia dette tante cattive parole			
5.			e io mi sono stato in solenzio		
6.			non ò risposto per niente		
7.				non creti come sto male	
8.				non mangio e non bevo	
9.				e se me vedi come sono ridotta	
10.				specialmente a desso che voi lo sai come sto male per la taranta	
11.					che non mi fa mangiare niente
12.					che morio della fame

1.	le mie vicine mianno detto che domandao di Maria se sto ancora come prima		
2.		e io dissi che nera potuto fare meno di questa cosa	

3.		che io non posso fare mai più pace di quello che a me mia fatto	
4.		che voi Mia Buona Signorina non puoi sapere tutto di quella da quantte mena fatte a me di tutti i colori	
5.			a quello che io lo fatto a Maria del bene
6.			che io li hò fatto da tanti anni
7.			che io doveva parare da tutto quello che lui voleva della mia casa

Da un punto di vista semantico, inoltre, la giustapposizione di questi elementi paradigmaticamente connessi, con chiaro effetto di ridondanza, offre un esempio di “plurideterminazione”, cioè del meccanismo di co-costruzione della referenza che rappresenta la strategia “più naturalmente compatibile con la comunicazione faccia a faccia” perché consente ai parlanti di procedere in maniera incrementale, mantenendo il testo prodotto coerente e coeso, nonostante le possibili sovrapposizione date da interventi esterni del destinatario, presente nello scambio e coinvolto nel dialogo (cf. Voghera 2017: 165). Nel caso di testo scritto, però, l’effetto ottenuto è il contrario: un testo prodotto in modalità scritta ma caratterizzato dalla pervasività di correlati funzionali tipici del parlato risulta scarsamente leggibile.

5.2 “[D]i animali e di leoni e di cavalli tutti maestrati”: *formulare e denotare attraverso le liste*

Secondo alcuni approcci, come quello proposto in Masini *et al.* (2018), strutture come quelle appena rappresentate nelle griglie possono essere considerate costruzioni a lista, di cui troviamo un uso pervasivo nelle lettere di Anna. A partire dagli anni Ottanta del Novecento, questo tipo di costruzione ha goduto di un’attenzione sempre crescente ed è stata indagata in campi-

oni rappresentativi di lingue, con diverse metodologie d'analisi e a partire da diverse cornici teoriche (fra le quali spiccano l'analisi conversazionale e la sintassi del parlato spontaneo). Secondo Masini *et al.* (2018: 50), possiamo definire costruzione a lista “the syntagmatic concatenation of two or more units of the same type (i.e. potentially paradigmatically connected) that are on a par with each other, thus filling one and the same slot within the larger construction they are part of”. Nell'ipotesi che le liste possano essere considerate come vere e proprie costruzioni nel quadro della *Construction Grammar*, presentando precisi tratti sia formali che semantici, le studiosse individuano una lista esaustiva di parametri formali che permettono l'individuazione di costruzioni a lista e un insieme di valori che possono essere veicolati da questo tipo di struttura.

Per quanto riguarda la struttura formale, Masini *et al.* (2018) osservano che gli elementi linguistici messi in lista possono essere di complessità variabile: si va da liste in cui si connettono paradigmaticamente singoli lessemi a liste di sintagmi di diversa natura o di frasi (nominali o verbali). Questi elementi possono contribuire a formare liste binarie (25), liste ternarie (26) o liste fondate su più di tre elementi paradigmaticamente connessi (27); è inoltre possibile distinguere, nella struttura formale delle liste, le costruzioni asindetichiche (28) da quelle sindetiche con proprietà congiuntive (29a), disgiuntive (29b) o avversative (29c). Ognuno dei tipi individuati ha esempi nelle *Lettere di Anna*¹⁰:

(25) LISTA BINARIA: ti penzzo sempre {tutti i momenti | ^ e li minuti}

(26) LISTA TERNARIA: ora timando {due frise | due taralli | ^ è questi noci}

(27) LISTA CON PIÙ DI TRE ELEMENTI: questo circo [...] era tutto {di animali | ^ e di leoni | ^ e di cavalli tutti maestri | ^ e tutto quello che hanno fatto prestigiatori delli animali}

(28) LISTA ASINDETICA: {Mia Buona Signorina | Mia Buona Mamma | Mio Buono Pappà | Mia Buona Signorina Paola}

¹⁰ Il tipo di rappresentazione delle liste adottato negli esempi è quello proposto da Masini *et al.* (2018: 57): «braces '{ }' delimit the list; the pipe sign '|' separates the conjuncts; round brackets '(')' indicate optionality; the circumflex accent '^' marks list markers, namely coordinators/connectives and list completers [...]».

- (29) a. LISTA SINDETTICA CONGIUNTIVA: tiò preparata una cosa buona che vuoi mangiate {colla vostra Madre | ^ è pure colla vostra sorella | ^ è pure il vostro Papà}
- b. LISTA SINDETTICA DISGIUNTIVA: che non sono {una studentte | ^ o una maestra}
- c. LISTA SINDETTICA AVVERSATIVA: tutte due ne abbiamo amate {come due Sorelle | ^ non come due amiche | ^ ma meglio di Sorelle }

Per quanto riguarda invece i valori semantici e funzionali veicolati dalle liste, Masini *et al.* (2018: 74) propongono la classificazione che si può osservare nello schema di riepilogo riportato in Figura 4.

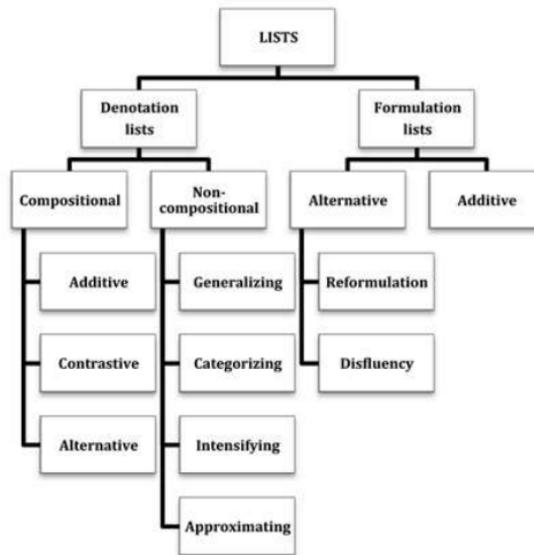


Figura 4. Classificazione semantica e funzionale delle costruzioni a lista in Masini *et al.* (2018: 74).

Nella tipologizzazione proposta dalle studiose vi è un’importante distinzione tra liste con funzione formulativa e liste con funzione denotativa. Le liste formulative sono quelle che coinvolgono più direttamente il livello discorsivo, si fondano su elementi paradigmaticamente connessi che rappresentano riformulazioni, ripetizioni e possono configurarsi come veri e propri esempi

della disfluenza testuale tipica del parlato. Come già osservato nel paragrafo precedente, i parlanti riformulano continuamente quanto affermano, sostituendo un costituente con un altro (si parla in questo caso di liste formulative alternative) o aggiungendo un costituente che esplicita o amplia quanto appena detto (si parla in questo caso di liste formulative additive). Le lettere dell'epistolario di Anna sono interamente strutturate come liste formulative, al punto che sarebbe difficile individuarne alcune specifiche: l'andamento sintattico del testo è interamente paratattico, fondato su giustapposizioni e ridondanze, come già mostrato negli esempi (21)-(24).

Più agevole è invece individuare, nel testo delle lettere, le liste denotative che servono a costruire nuove entità complesse, fondate sul riferimento alla denotazione di una serie di congiunti. Il significato delle nuove entità costruite attraverso una lista di congiunti è ricavabile dal significato degli elementi messi in lista e da quello dei marcatori di lista. Le liste di questo tipo possono essere distinte in composizionali e non-composizionali¹¹. Nelle liste con significato composizionale, gli elementi paradigmaticamente connessi possono essere rappresentati da lessemi, sintagmi o frasi che esprimono concetti strettamente connessi da un punto di vista semantico (e abbiamo in questi casi liste di co-ponimi, di sinonimi o di antonimi, come esempi di quella che Wälchli 2005 definisce *natural coordination*) o concetti che si trovano solo accidentalmente associati in una lista (e si parla in questi casi di *accidental coordination*, cf. Masini *et al.* 2018: 61). Negli esempi tratti delle lettere riportati di seguito abbiamo casi di liste denotative composizionali che rappresentano coordinazione naturale (30) o accidentale (31). In entrambi gli esempi, le liste contribuiscono a esplicitare un elemento introduttivo – *una bella mangiata* in (30) e *niente di bello* in (31) – definito da Bonvino *et al.* (2009) “pre-detailing”, che funziona “as a projecting «more-to-come» element, which is detailed and expanded by means of the list itself”.

(30) io e mia Sorella abbiamo fatta una giornata da legria e *una bella mangiata* {di pesce fatto alla salza | un po' di vino | un pezzo di pane | ^ e un piatto di pasta | ^ e

¹¹ La relazione semantica fra gli elementi in lista viene considerato da Masini *et al.* (2018: 61) come “a structural parameter, despite its inherent semantic nature, because what is here considered is the type of relation that characterizes the internal composition of the list set: in this sense, the semantic relation among the conjuncts functions as a structural constraint”.

pure opati pure due pazarotti di patate | ^ e per fruta ne abbiamo portato due mele grose) e quando abbiamo fenito di tutto o veduto Padre F. telo salotato (L2, p. 116)

(31) di quello che voi avete mandato alla tua Anna non cè proprio *niente di bello* {che la nello non mi va a nessuna parte | la catena non sono ragazza di portarla | la spilla non sono una ragazza piccola di portarla | il fazzo letto non melo posso proprio mettere {che tù lo sai che sono tarantata che mia noiono i colori | e non lo posso proprio guardare che mi sento troppo male} sala la bottiglia di profumo che tu losai che a me [+non] mi piace | i saponetti niopresi due scatole di quelli che passono in giro e no a veva bisogno} (L52, p. 173)

Se gli esempi in (30) e (31) si osservano all’interno delle griglie, è possibile osservare come nello scritto di Anna le liste possano, in alcuni casi, essere l’una incassata nell’altra. Il generale andamento paratattico e ridondante, caratterizzato da continue riformulazioni o – come nei casi appena osservati – ampliamenti tramite giustapposizione di elementi in lista può rappresentare un esempio di liste formulative che inglobano liste denotative, il cui significato compositazionale è in (30) rappresentato da elementi che entrano in una relazione semantica di co-iponimia (abbiamo una lista dei cibi che hanno costituito un pranzo) e in (31) da elementi coordinati accidentalmente, perché rappresentano il contenuto di un pacco regalo che la scrivente ha ricevuto¹².

1.	io e mia Sorella abbiamo fatta una giornata da legria	
2.	e una bella mangiata	
3.		di pesce fatto alla salza
4.		un po' di vino
5.		un pezzo di pane
6.		e un piatto di pasta
7.		e pure opati pure due pazarotti di patate
8.		e per fruta ne abbiamo portato due mele grose

¹² Va osservato che gli elementi in lista nell’esempio 30 (*anello, collana, spilla, foulard, profumo, saponette*) possono essere in qualche modo essere ricondotti a campi semantici affini (trattandosi di accessori o prodotti di cosmetica).

1.	di quello che voi avete mandato alla tua Anna			
2.		non c'è proprio niente di bello		
3.			che la nello non mi va a nessuna parte	
4.			la catena non sono ragazza di portarla	
5.			spilla non sono una ragazza piccola di portarla	
6.			il fazzo letto non melo posso proprio mettere	
7.				che t'ù lo sai che sono tarantata che mia noiono i colori
8.				e non lo posso proprio guardare che mi sento troppo male
9.			la bottiglia di profumo che tu losai che a me mi piace	
10.			i saponetti niopresi due scatole di quelli che passano in giro e no a veva bisogno	

Nelle liste non-composizionali, invece, il significato veicolato dagli elementi inseriti in lista può essere di tipo generalizzante (*generalizing*), categorizzante (*categorizing*), approssimante (*approximating*) o di intensificazione (*intensifying*), cf. Masini *et al.* (2018: 66).

La lista sindetica disgiuntiva in (29b), ad esempio, pare veicolare valore categorizzante: *non sono {una studentte | ^ o una maestra}* serve a creare una categoria vaga del tipo ‘persona istruita’. Lo stesso tipo di costruzione può essere rintracciata nell’esempio (32), in cui attraverso tre elementi in lista Anna crea per Annabella Rossi una categoria come ‘persona che ha qualcosa da perdere’, da contrapporre a quella cui sente di appartenere.

(32) di piu mi sono in teso il mio cuore mi batteva forte per te che diceva la tua
 Cara Anna che meglio di morire io che te perché io non faccio male a nessuno
 mentre tu siei {giovane | ^ e bella | ^ ed ai una famiglia}

Le formule di saluto che Anna usa in chiusura delle *Lettere* sono spesso liste asidentiche (ma non esclusivamente, come mostra l’esempio 33a) con valore di intensificazione, spesso veicolato dalla ripetizione di intere forme (32 b-e):

- (33) a. e ricevi {mille | ^ e mille {baci | ^ è bacioni}}
 b. ciao ciao ciao ciao ciao ciao ciao ciao baci baci baci baci a tutti BACI
 [L2, p.117]
 c. ciao ciao ciao ciao ciao Prontta risposta [L10, p.126]
 d. ciao ciao ciao sempre in attesa [L15, p. 133]
 e. Buone cose ciao ciao ciao ciao ciao [L16, p. 135]

5.3 “Mia Buona Cara Signorina”: *pragmaticalizzazione dell’allocutivo*

Una delle peculiarità della modalità parlata è senza dubbio la presenza di elementi polifunzionali che svolgono diversi ruoli interazionali o pragmatici, i cosiddetti segnali discorsivi¹³. Questi elementi, com’è noto, derivano principalmente da processi di pragmaticalizzazione di parole che originariamente erano nomi, verbi, avverbi, congiunzioni o intere clausole. La loro peculiarità sta nel fatto che non svolgono funzioni predicative o referenziali ma assolvono a una funzione eminentemente procedurale, aiutando i parlanti a “orientarsi” nell’interpretazione di un enunciato.

Nelle *Lettere* è possibile osservare l’evoluzione di una formula allocutiva con la quale Anna apre le sue lettere e che usa nel testo per rivolgersi alla destinataria: si tratta di sintagmi nominali che hanno come testa il nome *signorina*

¹³ Per una panoramica generale sui segnali discorsivi, cf. Sansò (2020)

(*Mia buona signorina, cara signorina, la mia buona signorina, mia signorina, Mia Buona cara signorina, buona signorina*) e che, sin dalla prima lettera, Anna utilizza frequentemente nel corso dei testi che scrive, come mostra anche l'analisi lessicale condotta con Voyant. Questa formula, nel corso dell'epistolario, inizia ad essere sempre più utilizzata con funzioni diverse da quella allocutiva: la multifunzionalità è, infatti, una delle caratteristiche cruciali dei segnali discorsivi¹⁴.

Oltre alla funzione allocutiva, sempre presente in apertura e che viene talvolta utilizzata anche nel corso del testo per riferirsi all'interlocutrice, i sintagmi con il nome *signorina* come testa possono veicolare anche valori metatestuali (in particolare, l'introduzione o la ripresa di un topic) e valori interazionali (il richiamo dell'attenzione), tipicamente veicolati dai segnali discorsivi¹⁵.

In tutto l'epistolario, abbiamo 332 occorrenze di sintagmi che hanno come testa *signorina*, nome utilizzato come titolo di cortesia da attribuire a donne nubili (quale Annabella Rossi, all'epoca dell'incontro con Anna, era): tutte queste occorrenze sono state annotate in base alla funzione svolta nelle lettere con le etichette **A** (funzione allocutiva), **M** (valore metatestuale), **I** (valore interazionale). Per esemplificare, in (34)-(35) è possibile leggere il testo delle lettere numerate rispettivamente come 38 e 59 nell'epistolario e osservare l'annotazione dei valori espressi dai sintagmi costruiti con *signorina*.

(34) Mia Buona Signorina[A]

dopo tanto tempo che desideravo un tuo caro scritto hò ricevuta la tua cartolina la quale sono rimata con tenta che hò saputo le tue notizie ma Mia Cara Signorina[A/I] la tua Anna non è mai sazia se non ricevo una bella lettera Mia Cara Signorina[M] ti farò sapere che il giorno di Pasqua lo fatto sempre piangendo per colpa tua che mi ai mandato un tuo scritto prima che tu lo sai il mio pensiero verso te che non sfugi un minuto della mia mente insoma Mia Cara Signorina[I] miai fatta in pazire non prendo riposo nenotte non giorno

14 Bazzanella (1995) distingue la multifunzionalità paradigmatica dalla multifunzionalità sintagmatica: in virtù della prima, un segnale discorsivo è in grado di veicolare funzioni diverse a seconda del contesto e del cotesto di occorrenza; in virtù della seconda, invece, lo stesso segnale discorsivo è in grado di veicolare, all'interno dello stesso contesto, diverse funzioni.

15 Come spiega Sansò (2020: 15-16), le tre funzioni principali attribuite ai segnali discorsivi nella letteratura sul tema sono quella interazionale, quella metalinguistica e quella cognitiva.

penzando te Mia Cara Signorina[I/M] vedendo che tardi a scrivere sospetavo una vostra improvvisata il giorno di S'amarco ma sono rimasta inlusa e io la festa lo pasata male perché pensavo che lanno scorso eri a casamia spero almeno di abbracciarti di S'anpaolo che ho tanto desiderio di vedere la vostra persona che a me e tanto Cara. Cara Signorina[M] quando mi tardi a scrivere il mio pensiero mi dice che forse tisiei sfogliata di scrivere alla tua Cara Anna e facevo come una pazza pregando alle mie vicine di casa di fare preghere di farmi avere notizie della Mia Anna bella Mia Buona Signorina[M] tutte le mattine quando andavo alla Santa Comunione la prima preghera era la Vostra pregavo prima di darti Salute secondo di dare notizie a me dunque spero che ai compreso il mio cuore e quello che tiracomando di scrivere presto e fami stare un po' tranquilla perché tu lo sai la mia malattia e sono calma soltanto quando vedo la tua posta che tengo un pensiero di una figlia che mi siei tanto tanto lontana che se fosti più vicina hò quante volte ti venivo a trovare e solo così mi potevo saziare Mia Cara Signorina[M] voglio sapere come avete passata la Santa Pasqua spero che state tutti bene in famiglia...

Mia Buona Signorina[M] un giorno era andata a un giardino cera una piccola ragazza che giocava con un pezzo di canna vecchia allor la presi io in mano quella canna e dal buco della Canna e uscito un scorpione grosso si afferro al braccio che mi pizzicò forte e mi senti tanto male che mi fece ballare come ballo a S'anpaolo per tante ore in maggina Cara Signorina[I] quello che ho passato come pure ti faro sapere che io incominciato a fare il primo giro in paese per quanto ariva la festa di Sanpaolo lotto maggio io parto per fare il giro ad altri paesi sempre per la festa di Sanpaolo.

ti preco tanto di mandare le foto dei Vostri genitori e della Vostra sorella Paola ricevi tanti saluti della Vostra Cara Anna e un saluto a signor G. e Sono la Serva Di Gesù

Pronta Risposta

(35) Mia Buona Signorina[A]

prima di tutto voglio sapere come state tutti di famiglia voglio sapere perché non ricevo più notizie di te che la tua Anna non sa quello che penza perche sia vicina il mese di S'ampaolo che io a spetava quella cosa che voi eri detto per una bella sopresa ma la tua Cara Anna non ha ricevuto proprio niente perché questa cosa e proprio tardata la risposta dalla Mia Buona Signorina[A] che ha spetto più la vostra bella lettera e che tutte le feste che vengono al mondo che io Cara Signorina[I] non tengo suna passione che vengono le feste la mia festa e la Vostra lettera che mi fa tanto piacere alla mia persona che nessuno tengo tanta cara come la persona della Mia buona Cara signorina[A] che io lo sò che mi ami di vero cuore non come le altre che dopo lo dato tutto quello che anno voluto mi-anno tradita con tutte le bastonate ma non li porta cè Dio, che se la paca coloro.

Mia Buona Signorina[M] se ti dico dime che tutti i santi giorni tengo quei tacchi e sempre penzo a voi che dico che non ti vedo più ma prima che la tua Anna andasse in Celo ti vorrei Vederti unaltra Volta se posso avere quelli onori. Mia Buona Cara Signorina...[M] ti dico che io sono andata angiro ai paesi per la lemosina di S'ampaolo per fare il pranzo delle Verginelle che tu lo sai come è di pacare le macchine che ci vole sei mila lire le sole macchine senza il pranzo che bisogna fare e senza tutte le altre cose e voi sapete quello che fa la tua Anna Mia Buona Signorina[V] come voi losapete che io vado angiro ai paesi e sono andata ma non ritrovata mia mipote e sono stata sola ma quando mi coricava alla sera mi sentiva lo recchio rusciare e io diceva mia scritto la Mia Buona Signorina[A] e non puoi mai cretere che gioia alla mia persona ma quando sono a rivata o perta la porta e non oveto proprio niente e mi sono venuti di più i nervi che li tengo... Mia Buona Cara Signorina[M] voglio sapere piagendo Dio. se vieni il giorno di S'ampaolo e voglio che mi mandi a dire prima di non fare le cose solite Vostre cose come ai fate le altre Volte e così se tu me lo dici io non mi ne esco di casa e di non andare anessuna parte solo se non mi trovi devi entrare a casa della Conziglia sola perché tu non sai niente che la Giovanna mia dette tante parolacce che tu non la sai come è che non sta più a dove abitava adesso sta a unaltra parte ma ne non mi teresa più di quella donna che io sto bene in casa mia per grazia di Dio, che non mi manca proprio niente... finisco dascrivere con le parole più belle che il Signere che ti ta una buona salute a tutti della Vostra famiglia e ricevi i più distinti Saluti della vostra Anna E Sono La Serva Di Gesù Pronta Risposta

Come si può osservare, la formula allocutiva viene rifunzionalizzata costantemente nel corso del testo. La multifunzionalità di questi sintagmi è, seguendo la proposta di Bazzanella (1995), sia paradigmatica che sintagmatica: la formula può, in altre parole, sia veicolare funzioni diverse a seconda del contesto e del cotesto di occorrenza (multifunzionalità paradigmatica) sia veicolare diverse funzioni nella stessa occorrenza, ad esempio quella allocutiva e quella interazionale di richiamo dell'attenzione (multifunzionalità sintagmatica).

Può essere interessante notare che, nel corso dell'epistolario, aumenta via via l'uso di questi sintagmi e, soprattutto, aumenta il loro utilizzo in funzioni assimilabili a quelle dei segnali discorsivi. Nella Tabella 5 questa osservazione è corroborata dai dati dell'annotazione di tre lettere, scritte in momenti diversi dell'epistolario.

Lettera 2 [12 occorrenze su 1782 totali]			Lettera 38 [10 occorrenze su 465 totali]			Lettera 50 [10 occorrenze su 681 totali]		
totale occorrenze sintagmi: 12			totale occorrenze sintagmi: 10			totale occorrenze sintagmi: 10		
Funzioni veicolate			Funzioni veicolate			Funzioni veicolate		
[A]	[I]	[M]	[A]	[I]	[M]	[A]	[I]	[M]
1	3	8	2	4	6	2	3	5

Tabella 5. Funzioni veicolate da SN con testa *signorina* in tre lettere del corpus.

5.4 “La tua Anna”: forme di vocativo e allocuzione inversa e di illeismo

Come osservato in 4.2, una delle parole più frequenti nell’intero corpus delle lettere è *Anna*. Questa massiccia presenza del nome della scrivente si spiega osservando l’uso di due diverse strategie discorsive, definite da Mazzoleni (1997) allocuzione inversa e vocativo inverso, ma a cui in letteratura si fa riferimento con etichette diverse, spesso sovrapposte. In alcuni lavori, infatti, con allocuzione inversa si fa riferimento a entrambi i fenomeni dei quali discutiamo o soltanto a quello che qui definiamo vocativo inverso (cf. fra gli altri Rohlfs 1925; Renzi 1968 e più recentemente De Vecchis 2022: 289-291). Per tenere distinti i fenomeni, in questo lavoro adottiamo le etichette di vocativo inverso (usato anche in Iovino & Rossi 2014) per identificare i casi in cui Anna nomina sé stessa in maniera slegata dalla sintassi della frase, di allocutivo inverso per quei casi in cui Anna si autoinserisce sintatticamente nella frase con ancoraggio deittico al destinatario, di illeismo per i casi in cui Anna parla di sé in terza persona. Tutti e tre i fenomeni sono molto tipici del *baby-talk* e del linguaggio familiare¹⁶.

Il vocativo inverso è una forma di allocuzione tipica di diverse varietà italo-romanze meridionali (e, per contatto, penetrata negli italiani regionali, cf. Sgroi 1990) in cui un locutore si rivolge al proprio interlocutore con una forma vocativa che identifica, però, sé stesso. In questi usi, le coppie di nomi di famiglia in relazione di inversione (come *mamma/*

¹⁶ Su queste due etichette, cf. almeno Moretti (2002) e Gordon (2009).

figlio, nonna/nipote, zia/nipote) appaiono “sostituite da un unico termine simmetrico che sembra assumere anche il valore del suo converso e neutralizzare l’opposizione” (cf. Mazzoleni 1997). Questa strategia occorre con maggior frequenza con verbi all’imperativo ma si usa anche in enunciati interrogativi o dichiarativi che, come nota Abbate (2010), possono veicolare tutte e sei le funzioni della comunicazione verbale secondo il modello di Jakobson. Da un punto di vista formale, le formule con funzione vocativa possono essere precedute dall’articolo (36a-c), in particolare in Sicilia o in Salento, o avere forma dativale (36d). Si osservino alcuni esempi tratti da Mazzoleni (1997).

- (36) a. Il padre al figlio: “Fa ‘o mare, *papà!*” (napoletano)
 b. Lo zio al nipote: “Va’, va’, zio, va’!” (italiano regionale campano, dal film *Scugnizzi* di Nanni Loi, 1989)
 c. La sorella al fratello: “*áni, la sòru*” (salentino, in Rohlfs 1966-1969, III, par. 657)
 d. Il nonno al/la nipote: “Mancia, *a nannò*” (siciliano, cit. in Varvaro 1988: 725)

Da un punto di vista pragmatico, nei diversi lavori che se ne sono occupati si sottolinea che questo costrutto ha la funzione di esprimere e rinforzare la solidarietà reciproca fra i membri di un gruppo familiare, evidenziando il rapporto di dominanza di un membro sull’altro. Il costrutto è infatti di tipo asimmetrico: può essere usato solo dall’alto della relazione familiare verso il basso ma non viceversa; attraverso questa strategia, quindi, si rimarca la propria superiorità, assicurando al contempo protezione al più giovane nella relazione¹⁷.

Nei casi di allocuzione inversa, invece, il mittente si inserisce nella frase con elementi che spesso lo ancorano al destinatario. Pragmaticamente il costrutto è affine al cosiddetto dativo etico che “indica la persona che partecipa emotivamente all’evento espresso nella frase” (Salvi 1989: 65), la quale

¹⁷ L’uso di strategie pragmatiche assimilabili al vocativo inverso è attestato in diverse lingue e varietà, con funzioni largamente sovrapponibili a quelle descritte per gli studi sulle varietà italo-romanze: si vedano ad esempio Ayoub (1964) sull’arabo; Boeder (1988) sul georgiano; Ferraro (1988) sul baluci.

entra nell’enunciazione anche se non è per forza beneficiaria dell’azione. Si notino gli esempi (37) di Mazzoleni (1997) e (38) di Gazzoni (2020):

(37) La madre al figlio: Portami i piatti, per favore

(38) La madre al figlio: Fammi i compiti, prima di uscire!

In entrambi i casi, oltre a mostrare la partecipazione del mittente all’enunciazione, l’effetto pragmatico dei costrutti risulta in una attenuazione della forza illocutoria dell’atto linguistico grazie a un maggiore avvicinamento emotivo.

Anche il costrutto che abbiamo definito illeismo è molto tipico del linguaggio familiare ma è stato osservato soprattutto in relazione ad altre forme di discorso (pratiche discorsive pubbliche¹⁸, patologie che investono il linguaggio¹⁹, funzione retorica in testi poetici²⁰). Una delle funzioni principali pare essere quella di animare la prospettiva dell’altro, attraverso un distanziamento. Questa strategia viene attuata dai parlanti che, nel riferirsi a sé stessi, utilizzano la terza persona singolare e tutte le forme che riguardano il polo opposto della deissi nell’enunciazione rispetto a quello del locutore. I parlanti, quindi, si auto-designano usando il proprio nome o, nel linguaggio familiare, il termine di parentela con il quale vengono designati dagli altri membri della famiglia. Sono forme di illeismo nel linguaggio familiare quelle in (39); nel discorso pubblico, un esempio di illeismo è in (40), estratto da un video pubblicato sulle proprie pagine social dall’ex senatore della Repubblica Nicola Morra. Questa strategia, presente in maniera consistente anche nel linguaggio poetico può essere usata anche in altri contesti, come l’intervista (in 41 si riporta una dichiarazione del calciatore Mario Balotelli) o il post sui social (in 42 si legge un estratto da un post su Facebook scritto dal cantante Povia).

(39) a. I genitori al figlio: “Tesoro, *mamma* e *papà* tornano subito”

b. La mamma al figlio: “*Mamma* ora si arrabbia!”

c. La nonna al nipote: “*La nonna* è stanca”

¹⁸ Questa strategia è largamente presente nei discorsi di Donald Trump, cf. Chilton (2017).

¹⁹ Cf. su questo Sterponi *et al.* (2015). Può essere inoltre indicativo notare come in un appunto dello psichiatra Giovanni Jervis contenuto tra le carte d’archivio di De Martino Anna viene descritta come affetta da crisi epilettiche frequenti, con facies e comportamento tipicamente maniacali, per le quali viene curata tramite del Talofen in gocce, un antipsicotico somministrato per il trattamento della schizofrenia e di altri disturbi di tipo psicotico (Signorelli & Panza 2011: 228).

²⁰ Per la presenza consistente di questa strategia nel linguaggio poetico, cf. Alek *et al.* (2020).

(40) Morra ha sempre avuto parole di rispetto verso chi è malato... Ma se voti Morra non ti devi lamentare se Morra provoca polemiche, perché Morra è così

(41) Mario è stato provocato e ha sbagliato a reagire. Ma non ha ammazzato nessuno

(42) Povia non è alternativo, è solo più attento e meno comodo di prima

Nelle sue lettere, Anna usa sia forme di vocativo e allocutivo inverso che di illeismo. Quest'ultima strategia è molto ben documentata nelle lettere come si può osservare già negli esempi in (12)-(13), o in quello in (43).

(43) con zidera la Vostra Cara Anna quantto a sofferto di tutte le mali sofferenze [L3, p. 120]

Anna usa, invece, una strategia che ricorda il vocativo inverso quando si rivolge alla sua interlocutrice per richiamarne l'attenzione auto-nominandosi, in una formula allocutiva introdotta dall'articolo determinativo che, come notava già Rohlfs (1966-1969, III, par. 657), è abbastanza comune in area salentina (e come osserva Sgroi 1990 anche in Sicilia).

(44) ma mi trovo a queste mie mali condizioni e non posso fare più niente che non è mia la mia possibilità per venire a trovarti la tua Anna [L62, p. 182]

In casi come quello in (45), le due strategie sembrano usate da Anna nello stesso contesto a breve distanza, nella prima occorrenza per introdurre sé stessa nel discorso usando la terza persona, nella seconda per richiamare l'attenzione dell'interlocutrice.

(45) la vostra Anna non era andata a nesuna parte cosi deve essere a visarla la tua Anna [L17, p. 136]

In altri casi, è invece difficile determinare se Anna usi una forma di vocativo inverso o se invece dia vita a un cambio di progetto che dà origine a un anacoluti.

(46) hò passato tutta la mia vita da quando era letà di venti sei anni le tà di sesantta tre anni in che moto li sto facendo che voglio mangiare è non posso mangiare per il fatto della tarantta perché voi losapete che Anna non mi manca propio niente di quello che io voglio ma la prendo perla more di Dio [L3, p. 117]

(47) Mia Buona Signorina ti mando a dire che la vostra Anna mi sento un po' più meglio [L18, p. 137]

Dal punto di vista dell'enunciazione del sé le lettere di Anna sono pertanto peculiari, perché permettono di osservare un *continuum* in cui il sé entra nella sfera dell'enunciazione a diverso titolo, a maggior ragione nei casi in cui è spesso difficile categorizzare in maniera univoca i costrutti in questione. Attraverso la menzione a sé stessa (sia attraverso casi di illeismo, che di allocutivo o vocativo inverso) Anna sceglie queste forme di auto-riferimento per sottolineare il suo ruolo di interlocutrice, mediando così la relazione in assenza con Annabella Rossi (Land & Kitinger 2007). Anna decide così di ancorare saldamente i suoi messaggi alla destinataria con mezzi deittici (come si può osservare dall'uso ripetuto di formule come *la tua Anna, la vostra buona Anna*), mantenendo al contempo la presenza di sé in quanto emittente. Questi mezzi linguistici che Anna mette in campo sembrano sia usati per ridurre la forza illocutoria delle continue e insistenti richieste, come l'invio di lettere, fotografie o regali, sia per sottolineare la sua partecipazione emotiva, apponendo, come nei casi di vocativo inverso, “una specie di firma [...] per rafforzare il particolare rapporto di intimità affettuosa” (Cardona 1987: 110).

6. Conclusioni

L'analisi delle *Lettere da un tarantata* proposta in questo lavoro ha permesso, attraverso un'analisi lessicale di tipo quantitativo e qualitativo, di osservare in primo luogo la conquista e l'evoluzione della scrittura nell'autrice delle lettere. Grazie a questo tipo di analisi, è possibile osservare le pratiche di alfabetizzazione nel loro divenire e di analizzarle come pratiche culturalmente situate. Seppur dotata di scarsa competenza linguistica, l'autrice delle *Lettere da una tarantata* è infatti in grado di utilizzare i pochi elementi a sua disposizione per una comunicazione di tipo emotivo, ossia intenzionalmente e strategicamente volta a influenzare la situazione e il rapporto con la destinataria (Caffi 2001). L'utilizzo esteso di specifiche formule linguistiche, come l'illeismo o il vocativo inverso, pare essere anzi un elemento che Anna gioca a proprio vantaggio per negoziare la presenza della destinataria, ossia Annabella, e per garantire la sottoscrizione e la continuazione del patto episto-

lare. Il parlare di sé pare perciò essere interpretabile come una vera e propria strategia di *footing* (cf. Goffman 1979), grazie alla quale Anna cambia cornice e anima le epistole evocando la presenza-assenza di Annabella, la quale non è solo assente fisicamente, ma anche emotivamente, come si intuisce dalle continue richieste di lettere da parte di Anna e dai contenuti dell'epistolario (che si configura come la testimonianza di un amore non corrisposto, cf. su questo Apolito 2015).

Inoltre, l'analisi dei principali fenomeni discorsivi individuati nelle *Lettere* ha permesso di osservare caratteristiche linguistiche e testuali che deviano dalle norme che regolano i testi prodotti in modalità scritta e che hanno come naturale modalità di ricezione la lettura e che costituiscono, al contempo, veri e propri correlati funzionali della modalità parlata.

Bibliografia

- Abbate, Lucia, 2010, "L'allocuzione inversa nell'italiano meridionale: una chiave interpretativa in base ai modelli pragmatici e cibernetici della comunicazione linguistica", in Pettorino, M., Giannini, A. & Dovetto, F. (eds.), *La comunicazione parlata 3, Atti del congresso internazionale (Napoli 23-25 febbraio 2009)*, vol I, Università degli Studi di Napoli L'Orientale, 145-160.
- Alek, Alek, Marzuki, Abdul G., Hidayat, Didin N., Islamiati, Farhana A., & Raharjo, Aning R., 2020, "'Why She Disappeared' (A Study of Illeism in Poetic Discourse)", *Ethical Lingua: Journal of Language Teaching and Literature* 7(2), 447-453.
- Allen, Barbara, 2015, "A non-standard standard? Exploring the evidence from nineteenth-century vernacular letters and diaries", in Auer, A., Schreier, D. & Watts, R. J. (eds.), *Letter Writing and Language Change*, Cambridge, University Press, 202-222.
- Apolito, Paolo, 2015, "E sono rimasta come lisolo a mezzo a mare", in Rossi, A. (ed.), *Lettere da una tarantata*, Squilibri, Roma, 13-62.
- Ayoub, Millicent R., 1964, "Bi-polarity in Arabic kinship terms", in Lunt, H. G. (ed.), *Proceedings of the Ninth International Congress of Linguists*, The Hague, Mouton, 1100-1106.
- Baig, Fatima Zafar & Naveed, Ahmed, 2019, "Literacy, Identity and Gender: A Case Study of Love Letter Writing Practices from Pakistan", *International Journal of English Linguistics* 9 (4), 288-306.
- Bazzanella, Carla (1995), "I segnali discorsivi", in Renzi, L. Salvi, G. & Cardinaletti, A. (eds.), *Grande grammatica italiana di consultazione*, vol. III (*Tipi di frase, deissi, formazioni delle parole*), Bologna, Il Mulino, 225-257.

- Bednarek, Monika, 2008, “Analyzing language and emotion”, in Bednarek, M. (eds.), *Emotion talk across corpora*, Basingstoke, Palgrave Macmillan, 1-26.
- Berruto, Gaetano, 2014, “Esiste ancora l’italiano popolare? Una rivisitazione”, in Daller, P. & Konecny C. (eds.), *Dall’architettura della lingua italiana all’architettura linguistica dell’Italia. Saggi in omaggio a Heidi Siller-Runggaldier*, Berna, Peter Lang, 277-290.
- Blanche-Benveniste, Claire, 1990, “Un modèle d’analyse syntaxique «en grilles» pour les productions orales”, *Anuario de Psicología* 47, 11-28.
- Blanche-Benveniste, Claire, Borel, Bernard, Deulofeu, José, Durand, Jacky, Giacomi, Alain, Loufrani, Claude, Meziane, Boudjema & Pazery, Nelly, 1979, “Des grilles pour le français parlé”, *Recherches sur le français parlé* 2, 163-206.
- Boeder, Winfried, 1988, “Über einige Anredeformen im Kaukasus”, *Georgika* 11, 11-20.
- Bonvino, Elisabetta, Masini, Francesca & Pietrandrea, Paola, 2009, “List Constructions: a semantic network”, Paper presented at the 3rd International AFLiCO Conference – *Grammars in Construction(s)*, Paris, 27–29 May 2009.
- Borges, M. J., 2016, “For the good of the family: migratory strategies and affective language in Portuguese migrant letters, 1870s–1920s”, *The History of the Family* 21 (3), 368-397.
- Burns, Tom R. & Carson, Marcus, 2005, “Social order and disorder: institutions, policy paradigms and discourses – an interdisciplinary approach”, in Wodak, R., Chilton, P. (eds.), *A New Agenda in Critical Discourse Analysis: Theory and Interdisciplinarity*, Amsterdam & Philadelphia, John Benjamins, 283-310.
- Caffarena, Fabio, 2005, *Lettere dalla Grande Guerra. Scritture del quotidiano, monumenti della memoria, fonti per la storia. Il caso italiano*, Milano, Unicopli.
- Caffi, Claudia, 2001, *La mitigazione. Un approccio pragmatico alla comunicazione nei contesti terapeutici*, Münster. Lit.
- Calaresu, Emilia, 2022, *La dialogicità nei testi scritti. Tracce e segnali dell’interazione tra autore e lettore*, Pisa, Pacini.
- Calaresu, Emilia, 2004, *Testuali parole. La dimensione pragmatica e testuale del discorso riportato*, Milano, FrancoAngeli.
- Cancian, Sonia, 2011, “Love in the Time of Migration Lovers’ Correspondence between Italy and Canada, 1948-1957”, *Diversité urbaine* 10 (2), 91-109.
- Cancian, Sonia, 2012, “The Language of Gender in Lovers’ Correspondence, 1946–1949”, *Gender & History* 24 (3), 755–765.
- Cardona, Giorgio Raimondo, 1987, *Introduzione alla sociolinguistica*, Torino, Loescher.
- Chilton, Paul A., 2017, “Toward a neuro-cognitive model of socio-political discourse, and an application to the populist discourse of Donald Trump”, *Langage et société* 160-161 (2), 237-249.

- Cortelazzo, Michele, 1972, *Avviamento critico allo studio della dialettologia italiana*, vol. III: *Lineamenti di italiano popolare*, Pisa, Pacini.
- D'Achille, Paolo, 2002, *Italiano dei semicolti e italiano regionale. Tra diastratia e diatopia*, Limena (PD), Libreria Universitaria.
- De Mauro, Tullio, 1970, "Per lo studio dell'italiano popolare unitario", in Rossi A., *Lettere da una tarantata*, Bari, De Donato, 43-75.
- De Vecchis, Kevin, 2022, *Il romanesco periferico: un'indagine sul campo*, Pisa, Pacini.
- D'Onghia, Luca, 2014, "Michelangelo in prosa: sulla lingua del «Carteggio» e dei «Ricordi»", *Nuova Rivista di Letteratura Italiana* XVII (2), 89-113.
- Duranti, Alessandro, 2005, "Ethnography of Speaking: Toward a Linguistics of the Praxis", in Kiesling, S. F. & Paulston, C. B. (eds.), *Intercultural Discourse and Communication: The Essential Readings*, Malden Mass, Blackwell, 17-33.
- Fairclough, Norman, 2001, *Language and Power* (2nd edition), Longman, London.
- Fairclough, Norman; Wodak, Ruth, 1997, "Critical Discourse Analysis", in van Dijk, T. A. (eds.), *Discourse Studies: A Multidisciplinary Introduction* vol. 2, Sage, London, 258-284.
- Ferraro, Felicetta, 1988, "Baluchi Kinship Terminology", *Newsletter of Baluchistan Studies* 5, 33-62.
- Fresu, Rita, 2015, "Scritture e Grande guerra: una storia linguistica tra 'alti' e 'bassi'", in Fresu, R. (ed.), *Questa guerra non è mica la guerra mia: scritture, contesti, linguaggi durante la Grande Guerra*, Roma, Il Cubo, 7-31.
- Fresu, Rita, 2020, "Dalla devianza al continuum. L'italiano dei semicolti negli studi storico-linguistici: evoluzioni e linee di tendenza", in Lobin, A., Dessì Schmid, S., Fesenmeier, L. (EDS.), *Norm und Hybridität / Ibridità e norma. Linguistische Perspektiven / Prospettive linguistiche*, Berlin, Frank & Timme, 249-267.
- Gazzoni, Sandra, 2020, *Corpus in fabula. Imperativi benefattivi come azioni congiunte*, Tesi di dottorato, Università di Bologna.
- Gee, James Paul, 2015, "The New Literacy Studies", in Rowsell, J. & Pahl, K. (eds.), *The Routledge Handbook of Literacy Studies*, London, Routledge, 35-48.
- Gerber, David A., 2005, "Acts of deceiving and withholding in immigrant letters: Personal identity and self-presentation in personal correspondence", *Journal of Social History* 39 (2), 315-330.
- Gerber, David A., 2006, *Authors of their lives: the personal correspondence of British immigrants to North America in the nineteenth century*, New York, University Press.
- Goffman, Erving, 1979, "Footing", *Semiotica*, 25 (1-2), 1-30.
- Gordon, Cynthia, 2009, *Making Meanings, Creating Family*, Oxford, Oxford University Press.

- Guidolin, Gaia, 2011, *Analisi linguistica del carteggio di Pietro e Alessandro Verri (1766-1797)*, Tesi di dottorato, Università di Padova.
- Heldner, Mattias & Hedlund, Jens, 2010, “Pauses, gaps and overlaps in conversations”, *Journal of Phonetics* 38 (4), 555-568.
- Land, Victoria & Kitzinger, Celia, 2007, “Referring to self and others in conversation”, *Discourse Studies*, 9 (4), 493-525.
- Lerner, Gene H., 1994, “Responsive List Construction: A Conversational Resource for Accomplishing Multifaceted Social Action”, *Journal of Language and Social Psychology* 13 (1), 20-33.
- Lucisano, Pietro & Piemontese, Maria Emanuela, 1988, “Gulpease. Una formula per la predizione della difficoltà dei testi in lingua italiana”, *Scuola e Città* 3, 110-124.
- Masini, Francesca & Pietrandrea, Paola, 2010, “Magari”, *Cognitive Linguistics* 21 (1), 75-121.
- Masini, Francesca, Mauri, Caterina & Pietrandrea, Paola, 2018, “List constructions: Towards a unified account”, *Italian Journal of Linguistics*, 30 (1), 49-94.
- Mazzoleni, Marco, 1997, “Vocativo, termini di parentela, e vocativo inverso: fra valori di sistema e valori di campo”, in Agostiniani, L., Bonucci, P., Gianecchini, G., Lorenzi, F. & Reali, L. (eds.), *Atti del Terzo Convegno della Società Internazionale di Linguistica e Filologia Italiana (Perugia, 27-29 giugno 1994)*, Napoli, Edizioni Scientifiche Italiane, 457-475.
- Mengaldo, Pier Vincenzo, 1987, *L'epistolario di Nievo: un'analisi linguistica*, Bologna, Il Mulino.
- Moretti, Bruno, 2002, “Affettività nel baby talk”, in Bazzanella, C. & Kobau, P. (eds.), *Passioni, emozioni, affetti. Dynamie*, Milano, McGraw-Hill, 279-289.
- Nodari, Rosalba & Corona, Luisa, in revisione, “Un gioco di sguardi che non si incontrano: scritture marginali e autorialità sovrapposte in Lettere da una tarantata”.
- Renzi, Lorenzo, 1968, “Mamă, tată, nene, ecc. il sistema delle allocuzioni inverse in rumeno”, *Cultura neolatina* 28 (1), 89-99.
- Rohlf, Gerhard, 1925, “Un problema di sintassi italiano-meridionale (abruzz. ‘mam-masé’ = figliuol mio)”, *Archivum Romanicum* 9, 439-443.
- Rohlf, Gerhard, 1966-69, *Grammatica storica della lingua italiana e dei suoi dialetti* (trad. it.), Torino, Einaudi.
- Rossi Annabella (2015 [1970]¹), *Lettere da una tarantata*, Roma, Squilibri.
- Rovere, Giovanni, 1977, *Testi di italiano popolare: autobiografie di lavoratori e figli di lavoratori emigrati. Analisi sociolinguistica*, Roma, Centro studi emigrazione.
- Salvi, Giampaolo, 1989, “La frase semplice”, in Renzi, L., Salvi, G. & Cardinaletti, A. (eds.), *Grande grammatica italiana di consultazione. vol. 1*, Bologna, Il Mulino, 29-114.

- Sanga, Glauco, 2011, "Lettere da una tarantata di Annabella Rossi (1970) lette da Glauco Sanga", in Casellato, A. & Levis Sullam, S. (a cura di), *Leggere l'unità d'Italia: Per una biblioteca del 150°*, Venezia, Edizioni Ca' Foscari, 98-102.
- Sansò, Andrea, 2020, *I segnali discorsivi*, Roma, Carocci.
- Sgroi, Salvatore C., 1990, "Sull'interpretazione dell'allocuzione inversa", *Bollettino del Centro di Studi Filologici e Linguistici Siciliani* 16, 240-249.
- Signorelli, Amalia & Panza, Vittorio (eds.), 2011, *Ernesto De Martino. Etnografia del tarantismo pugliese: i materiali della spedizione in Salento del 1959*, Lecce, Argo.
- Sinclair, Stéfán & Rockwell, Geoffrey, 2012, The Voyant Tools Team, Voyant Tools (web application: 2012). <http://voyant-tools.org/?view=CollocatesGraph> [accesso in data 23/02/2023].
- Sornicola, Rosanna, 1981, *Sul parlato*, Bologna, Il Mulino.
- Spitzer, Leo, 1976, *Lettere di prigionieri di guerra italiani, 1915-1918*, Torino, Bollati Boringhieri.
- Sterponi, Laura, de Kirby, Kenton & Shankey, Jennifer, 2015, "Subjectivity in autistic language: Insights on pronoun atypicality from three case studies", in O'Reilly, M. & Lester, J. N. (eds.), *The Palgrave handbook of child mental health*, New York, Palgrave Macmillan, 272-295.
- Vanelli, Laura, 2008, "Italiano popolare e dialetti in un epistolario friulano della I Guerra Mondiale", in Marcato, G. (ed.), *Dialetto. Uso. funzioni. forma. Atti del convegno Sappada/Plodn, Belluno, 25-29 giugno 2008*, Padova, Unipress, 161-170.
- Varvaro, Alberto. 1988, "Italiano, Aree linguistiche XII: Sicilia", in Holtus, G., Metzeltin, M. & Schmitt, C. (Hrsgg.), *Lexikon der Romanistischen Linguistik, IV: Italienisch, Korsisch, Sardisch*, Tübingen, Niemeyer, 716-31.
- Vitali, Giovanni Pietro, 2020, What is a last letter? A linguistic/preventive analysis of prisoner letters from the two World Wars, in Marras, C., Passarotti, M., Franzini, G. & Litta, E. (eds.), *Atti del IX Convegno Annuale AIUCD. La svolta inevitabile: sfide e prospettive per l'Informatica Umanistica*, in *Quaderni di umanistica digitale*, 265-272.
- Voghera, Miriam, 2017, *Dal parlato alla grammatica. Costruzione e forma dei testi spontanei*, Roma, Carocci.
- Volpi, Mirko, 2010, "«Sua Maestà è una pornografia!». La scrittura della protesta nelle lettere al Re durante la Grande Guerra", *La lingua italiana* VI, 123-140.
- Wälchli, Bernard 2005. *Co-compounds and natural coordination*. Oxford: Oxford University Press.

FRANCESCA D'ANGELO*

Assessing translation quality: a survey of research into human translation, post-editing and machine translation

Abstract

L'articolo analizza il rapporto tra traduzione umana, post-editing e traduzione automatica affrontando, in particolare, la discussa questione della qualità traduttiva. Vengono presentati diversi approcci e metodologie, identificando una serie di problematiche riguardanti la traduzione in termini di processo, prodotto e sistemi di valutazione. È inclusa, inoltre, una riflessione sull'osservata tendenza all'oggettivazione degli standard di traduzione e all'utilizzo di metriche automatiche, con particolare attenzione alle possibili implicazioni che ciò potrebbe avere per la comunità scientifica e di traduzione.

Parole chiave: Traduzione automatica; Post-Editing; Traduzione umana; Qualità traduttiva

The article investigates the relationship between human translation, post-editing and machine translation addressing, in particular, the controversial issue of translation quality. An overview of different methodologies and approaches is provided, identifying a number of perceived issues concerning translation in terms of process, product, and evaluation systems. A discussion on the observed tendency to objectify translation standards, and to automatic metrics, is included with a focus on the implications it may have for the research and translation community.

Keywords: Automatic Translation; Post-Editing; Human Translation; Translation Quality

1 Introduction

The considerable growth of interest in the translation field in recent years, due to the pressure to increase the productivity of translation in terms of both amount of text and processing time, has led researchers from multiple fields of study including linguistics, computer sciences and artificial intelligence to investigate the role of machine translation (MT). On one hand, automat-

* Francesca D'Angelo, Alma Mater Studiorum Università di Bologna, francesca.dangelo16@unibo.it.

ic translation is seen as a challenging opportunity to foster and support the translation process, in the form of post-editing, increasing the speed and productivity of translators. On the other, the increasing spread of MT software, to respond to the significant growth of linguistic content to translate, contributes to questioning the translation output in terms of expected quality level. Starting with a technical overview of the most influential approaches to MT, from Statistical Machine Translation (SMT) to the advent of Neural Machine Translation (NMT) systems, the paper aims to provide a historical background of automatic translation and post-editing. In particular, the change of perspective towards post-editing, presented from a diachronic perspective from its introduction in the late 50s to the latest implementations in different fields, shows how the changing demand has brought an overall amelioration of the process. Besides, the article sheds light on the relationship between human translation and MT addressing the controversial issue of translation quality, human parity and superhuman performance, highlighting the complexity of the concept, its different categorisations as well as the different approaches available, from human to automatic quality evaluation systems. Finally, a debate on the increasing tendency to objectify translation quality, through the development of indicators and standards, is included with a focus on its impact in terms of needs and expectations concerning the quality assessment of MT and post-edited texts. In particular, the paper addresses the most recent developments, since 2013, regarding the role of post-editing effort in assessing the quality of neural machine-translated texts.

2 Approaches to Machine Translation: a Critical Overview

MT is a recently developed subfield of computational linguistics that inquires how computer software can develop efficient systems that can translate between human languages (Oladosu et al. 2016). Before starting the discussion on the productivity and quality of MT output, a technical overview of the most important approaches to MT is necessary to better understand the effectiveness of automatic translation and how it can support human translation. A first distinction can be made in MT between single and hybrid approaches to investigate their mechanism, affordances and constraints. Afterwards, NMT is presented as a new approach in the field of automatic machine translation. It has

gained great popularity among researchers in this field because of the promising translation results achieved, in addition to the simplicity of its structure.

Single approaches to MT encompass rule-based, direct-based, corpus-based, and knowledge-based methods. Rule-Based MT employs morphological, syntactic, and semantic rules to address word order issues. Direct-Based MT relies on bilingual dictionaries for word substitution, followed by syntax rearrangement, suitable for unidirectional translation. Corpus-based MT includes example-based and statistical methods, with the latter utilising machine learning on parallel corpora for translating new sentences. Finally, knowledge-based MT relies on extensive semantic and pragmatic knowledge for translation choices.

In contrast, the hybrid approach combines multiple MT techniques, typically integrating statistical and rule-based approaches. This hybrid approach includes word-based, phrase-based, syntax-based, and forest-based models. Word-based models focus on lexical word dependencies but struggle with long-distance reordering. Phrase-based models, on the other hand, use phrases as translation units, allowing local reordering and handling idiomatic expressions. Finally, syntax-based models analyse hierarchical sentence structures, offering string-based and tree-based parsing options. However, it should be noted that tree-based systems may suffer parsing errors due to relying on a single best parse tree. To overcome these issues, forest-based translation was introduced as a hybrid approach, blending string and tree-based methods to enhance MT efficiency and reduce parsing errors. It is a combination that facilitates faster decoding, making it a valuable addition to the MT landscape. In summary, the choice between single and hybrid MT approaches depends on the specific requirements of a translation task, with hybrid approaches offering advantages in terms of flexibility and quality improvement through the integration of multiple techniques.

These translation systems, however, still present significant flaws that more recent translation machine translation systems have tried to overcome. Indeed, past MT systems were mainly rule-based systems with the aim of creating grammatical rules for the source and target language. Indeed, MT acted as a translation between languages based on this set of rules. The problem was mainly the addition of new content and new language pairs

since maintaining and extending such a set of rules was too time-consuming and costly. Hence, SMT was created to respond to these translation problems (Koehn 2010). SMT systems create statistical models by analysing an aligned set of source and target language sentences (i.e. training set). On one hand, the advantage of SMT concerns its automatic learning process and the relatively easy adaptation. The disadvantage, on the other hand, refers to the training itself as it is necessary to create a usable tool and a large database of source and target language segments. Another critical aspect of SMT arises when dealing with grammatically more complicated languages.

In detail, the NMT has recently started to be promoted to solve these technical issues associated with SMT systems. To give some indication of the speed of change, in 2015, only one neural machine translation system was submitted at the shared task for machine translation organized by the Conference on Machine Translation (WMT). In 2017, almost all submitted machine translation systems were neural. The system looks at the sentence as a whole and can form associations between phrases even at greater distances in the sentence. In particular, at the IWSLT 2015 evaluation expedition, NMT was able to overcome state-of-the-art phrase-based machine translation systems on English-German, the language pair famous for its difficulty due to morphology and grammatical differences (Bentivogli et al. 2018).

The SMT system consists of several components tuned separately, whereas the NMT model is a large end-to-end single network that consists of two sub-recurrent neural networks: the encoder and the decoder. Moreover, if the SMT system needs many features that are accurately defined to do the translation, the NMT model depends on a training corpus to learn the translation task, with less or no feature engineering effort by linguists. Another major advancement that is worth discussing concerns the ability of NMT to seize potential long-distance dependencies and complicated word alignment information. In addition, the NMT model does not require a large memory space, such as those used by the SMT to store a translation model, a reordering model and a language model.

Nonetheless, it is important to notice that although convolutional and sequence-to-sequence provided good translation accuracy, the latter was reduced as the length of the input sentence increased. These models have

adopted an encoder–decoder approach that compresses all the necessary information of a source sentence into a fixed-length vector, which made it difficult for the models to handle long sentences, especially those longer than the training sentences. This problem has been solved with the introduction of the attention mechanism, which has achieved great popularity and has been used in various fields. In MT, the three architectures used are Stacked RNN with Attention, Self-attentional Transformer, and Fully Convolutional Models (ConvSeq2Seq). In 2015, to deal with the problem of fixed-length vectors, Bahdanau et al. (2014) proposed a model that extends the encoder–decoder approach by allowing automatic search for portions of a source sentence, which have relevance to the prediction of a target word, without explicitly forming these portions as a hard segment. Instead of encoding a whole input sentence into a vector of fixed length, the model converts it into a sequence of vectors. Each time during the decoding process, the decoder searches the input sentence for the words that have the most relevant information to generate the target word. The target word is predicted based on a context vector of all relevant words, and all previously predicted target words.

3 Background of Machine Translation and Post-Editing Research

The changing landscape in the translation industry, due to the recorded technological advances, has raised important questions on the relationship between human translators and MT. In particular, a change of perspective towards MT output has been acknowledged in the last decades. From its advent, an overall scepticism could be observed towards machine translated works and their uses in industries were peripheral and limited. Nowadays, thanks to the availability of free online MT systems, e.g. *DeepL*, *Bing Microsoft Translation*, *Reverso*, *Smartling* etc., translation output has started to be used as a rough version to be post-edited by human translators. More precisely, “post-editing is the correction of raw machine translated output by a human translator according to specific guidelines and quality criteria” (O’Brien 2001: 1). The advances recorded in the MT field, have brought the translation industry to increase the demand for post-editing instead of translating from scratch or revising human translations. If on one hand, the interest in the field is relatively recent, post-editing is one of the earliest uses envisioned for MT systems.

The history of post-editing can be dated back to the late '50s and early '60s when it was considered a “surprisingly hot topic” (García 2012: 293). One of the first officially recorded uses of post-editing in MT systems refers to the translation of scientific texts from Russian to English at the RAND Corporation (see Edmundson and Hays 1958). This approach entailed a good command of English on behalf of the post editor but not necessarily knowledge of the source language. Indeed, he/she would work on the MT text supported by a grammar code indicating the morphological information of the case, number, part of speech etc. In the 60s, post-editing was employed by the US Air Force's Foreign Technology Division and Euratom. However, after a negative report by the Automatic Language Processing Committee in 1966 pointing out that it was not worth the effort in terms of quality and productivity compared to human translation, post-editing systems stopped receiving funds (Koponen 2016).

Despite not living up to the expected performance in terms of time, quality and productivity, MT and post-editing continued to be developed and refined. Starting from the 70s, post-editing processes were implemented by important organisations such as the EU and the Pan-American Health Organisation. However, balancing between the advantages of post-editing in terms of speed and productivity of translators on one hand and the translation quality is still a controversial issue in translation research. An influential work by Gaspari et al. (2015) provides an interesting overview of the perception of MT texts and the use and effectiveness of post-editing in the translation and localisation field. The large-scale survey of MT competencies, based on data derived from 438 validated respondents including freelance translators, language service providers, translator trainers and academics, reveals interesting information about the needs and expectations of the community of translators.

The study highlights the increasing use of MT and post-editing, the increased demand for this service compared to the past and infers that this demand is expected to rise in the future. In more detail, the authors found that MT was used by 30% of the respondents, while 21% considered it useful and declared to be likely to use it in the future. Interestingly, 38% of the participants who declared to use MT pointed out that the text was always post-edited, 32% of them never performed post-editing and the remaining 32% of the interviewees used post-editing discontinuously. However, an important

observation to clarify the data referred to the overall level of satisfaction is needed. Most participants who declared to use MT did not use any customisation tool before performing the translation. Since half of the respondents show a low level of satisfaction, the authors analysed the correlation between the two factors finding that those who did not customise MT, were not satisfied with the translation. This explains how fundamental customisation is to meet the translation needs more specifically and to ameliorate the quality of the translation output. As Gaspari et al. (2015) suggest, it is possible to improve customisation by ameliorating the translators' assets and technological competencies such as customised glossaries in linguistic pre-processing of the text. Besides, the respondents who declared to perform post-editing in MT resorted to human evaluation as the most common approach to quality assessment, another controversial issue that is worth discussing in more detail.

An interesting study by Cettolo et al. (2013) provides valuable insights into assessing the quality of neural machine-translated texts, particularly focusing on post-editing efforts. Indeed, in 2013, a novel approach was introduced to evaluate machine translation output. This evaluation took place during the 10th IWSLT evaluation campaign, centred on transcribing and translating lectures using the TED Talks corpus. The assessment encompassed various language pairs, including English-German-French, with optional tracks for 12 languages. Eighteen teams participated, submitting 217 runs, which were evaluated using objective metrics and compared to previous systems. Instead of traditional subjective rankings, the study investigated the post-editing effort required by professional translators to improve machine-generated translations. This led to the adoption of the Human-mediated Translation Edit Rate (HTER) metric, which measures the minimal edit distance between the machine-generated translation and its manually revised version. HTER demonstrated a strong correlation with human evaluations of translation quality. Overall, the post-editing task offered a dual advantage: it highlighted specific translation errors and provided additional reference translations, enhancing the assessment of MT systems. Notably, the study revealed that the most proficient system required minimal post-editing effort, underscoring the potential of machine translation to assist human translators, with an optimal HTER score threshold of 19%.

As already argued, in 2015, a pivotal moment in MT occurred when a NMT system, as detailed by Luong and Manning (2015), surpassed Phrase-Based Machine Translation (PBMT) systems in the IWSLT competition. This marked a substantial improvement in translation quality, especially for complex language pairs like English-German, heralding the onset of the NMT era, following a period when NMT was computationally and resource-intensive compared to PBMT. For the purpose of the current discussion, it is worth recalling a work by Bentivogli et al. (2018) which sheds light on the role of post-editing in MT quality assessment, taking into account the developments of NMT quality. Particularly, the study underscores NMT's substantial advancements, showcasing its superior translation quality and post-editing efficiency across challenging language pairs and diverse sentence lengths. More specifically, the research compared NMT and PBMT outputs by analysing high-quality post-edits performed by professional translators on IWSLT data. This approach, unlike conventional MT evaluation reliant on arbitrary reference translations, enabled a comprehensive evaluation of systems, incorporating post-editing effort and error types. It also holds practical relevance for integrating MT into Computer-Assisted Translation (CAT) tools, where post-editing is common. The key findings highlighting NMT's superiority over PBMT can be summarised in the following points:

- NMT notably reduced overall post-editing effort.
- NMT consistently outperformed PBMT across various sentence lengths.
- NMT generated a higher proportion of low-error MT outputs, crucial for CAT tools.
- NMT exhibited significantly fewer errors, with lower error rates.
- NMT produced fewer lexical, morphology, and word order errors compared to PBMT.

4 Translation Quality: from Human Evaluation to Automatic Metrics

What clearly emerges from the current discussion about post-editing is that the task significantly differs from the traditional process of translation and revision. Considering how common this practice has become, organisations implementing MT are now trying to balance cost and productivity to assess translation. Nonetheless, getting to commonly shared metrics to as-

sess translation quality is not an easy task since the definition of translation quality may significantly vary depending on the factor under investigation.

Indeed, studies addressing the quality of MT texts have been concerned about the supposed inferior level of MT texts compared with the quality level of manually translated texts. A study by Fiederer and O'Brien (2009) compared a set of sentences translated manually or by post-editing to address this question. The different sentences were rated according to three criteria: i.e. clarity (how understandable the sentence was); accuracy (how close the target text meaning was to the source text); and style (naturalness and appropriateness). The findings indicate that post-edited translations were rated higher in terms of clarity and accuracy whereas, in terms of style, the version of manually translated texts was preferred. On the other hand, Carl et al. (2011) asked evaluators to rank in order of preference the manually translated texts and the post-edited versions. A slightly higher, although not significant, preference was recorded for the post-edited translations. However, to better interpret these findings, it would be worth including other variables in the analysis such as the proficiency of the evaluators in both languages, experience in the translation field etc. Other studies assessing the quality of MT texts have focused on the number of errors found in the translation. For instance, Plitt and Masselot (2010), assessed manually and post-edited versions of translated texts according to the criteria employed by the company's quality assurance team. Interestingly, although both versions were considered acceptable for publication, the evaluators considered the manually translated texts as needing more corrections.

On the whole, the literature on MT and human translation evaluation shows that post-editing can lead to quality levels close to manually translated texts. Depending on the quality criteria employed, post-editing texts are, in fact, sometimes even preferred to the manual versions. Nonetheless, all these studies discussed and reviewed so far involve human evaluators to rate in order of preference or choose between manually translated and MT versions. Currently, the trend has moved from human evaluation towards automatic tools since they are less time and cost-consuming. It should be noticed, indeed, that human judgement is mainly based on two main criteria, i.e. adequacy and fluency, and it is rather subjective. Hence, automatic eval-

uation measures are sought. The most common and better developed nowadays are BLEU (Bilingual Evaluation Understudy); WER (Word Order Rate); PER (Position-Independent Word Error Rate); and NIST (National Institute of Standards and Technology). Each automatic evaluator is developed on certain standards and has its affordances and constraints.

Developed by Papineni et al. (2002), BLEU addresses the evaluation problem by comparing the system output with a reference translation of the same text. The validity of BLEU has been proved through correlations with human evaluations. WER is developed by computing the number of substitutions, insertion and deletion operations performed to convert the generated translation into the reference translation. Where several reference translations are provided as source text, the evaluator calculates the minimal distance to this set of references (Nieben et al. 2000). One of the main disadvantages of WER is that it requires perfect word order. To solve this issue, the position-independent word was introduced (i.e. PER). NIST, on the other hand, is based on the BLEU metrics but with some differences. It calculates the *n*-gram precision the same way as BLEU but, at the same time, it also calculates how informative a particular *n*-gram is. As O'Brien (2011: 3) points out, the limitations of these automatic metrics are well acknowledged by the translation community. Indeed, automatic metrics are not supposed to properly predict the usefulness, adequacy, and reliability of MT technologies. In addition, it can be argued that if on one hand the usefulness of automatic evaluation metrics is deemed, on the other, they believe that too much importance has been given to them "since real translation quality is what we should be concerned with" (O'Brien 2011: 3).

Besides, on the expected level of translation quality, it is important to notice that the discussions surrounding the achievements of 'human parity' and 'super-human performance' in the domain of Neural Machine Translation, as asserted respectively by Hassan et al. (2018) and Bojar et al. (2018), have engendered substantial scholarly scrutiny. These assertions have catalysed an extensive examination of the appropriateness of the evaluation metrics employed in the assessment of NMT systems. Presently, the landscape of neural methodologies in machine translation presents novel challenges, with NMT outputs exhibiting a remarkable fluency. It has become evident that traditional automated metrics may inadequately capture the nuanced quali-

ty of neural systems, which are distinguished by their elevated fluency. This recognition underscores the imperative for the development and adoption of more nuanced and context-sensitive evaluation approaches, as witnessed by different researchers (e.g. Belouadi & Egere 2022; Mathur et al. 2020; Marie et al. 2021), to comprehensively assess the capabilities and limitations of these advanced neural machine translation systems. The aforementioned terminology such as “human parity” (Hassan et al. 2018) and even “super-human performance” (Bojar et al. 2018; Barrault et al. 2019) has been employed, particularly within the context of the Workshop on Machine Translation (WMT) evaluation campaigns. This suggests that MT systems are presumed to have reached a level of quality equal to, or possibly surpassing, the level of human translation, at least in the specific evaluation framework employed. In practical evaluations, machine-generated translations were consistently preferred over those produced by professional human translators. While these achievements appear impressive on the surface, it is imperative to engage in a more extensive and nuanced discussion, placing these claims into a broader context for comprehensive analysis.

A remarkable study by Hassan et al. (2018) investigates the challenge of defining and accurately evaluating human parity in translation. The authors adopt the following definition: “If there is no statistically significant difference between human quality scores for a test set of candidate translations from a machine translation system and the scores for the corresponding human translations then the machine has achieved human parity” (Hassan et al. 2018: 2). The paper provides an overview of Microsoft’s machine translation system and assesses the translation quality on the well-established WMT 2017 news translation task, specifically from Chinese to English. The findings indicate that Microsoft’s latest neural machine translation system has achieved a new state-of-the-art performance, and its translation quality is on par with that of professional human translations. Moreover, it significantly surpasses the quality of translations produced by non-professional crowd-sourced sources.

The same study has been reassessed by other groups of researchers, such as Toral et al. (2018), who observed specific shortcomings referred to the WMT evaluation. First, the translations to be evaluated were problematic

since the source texts were translations from other languages. That is, texts which were not considered appropriate for evaluation since they may present issues with paraphrasing, idiomaticity etc. Second, the evaluators were not always chosen among professional translators but they included participants of the study or remote crowd-workers, preferring more direct translations. Another critical aspect observed by the authors concerns the perception of translation quality, which presents a margin of variability according to the evaluator (i.e. end-users, MT developers, and professional translators). Similar conclusions were propounded by a work by Läubli et al. (2020). More specifically, the experiment's results pointed to problems dealing with the type of evaluation performed at the 2018 WMT by text segment. Hence, it cannot really take into account the text as a coherent whole. In addition, as argued by the authors, the perceived quality in human evaluation depends on a number of variables such as the choice of the evaluators, the availability of the linguistic contexts and the creation of reference translations.

On the controversial notions of human parity and superhuman performance in MT studies, it is worth recalling Toral's (2020) point of view. In his work, the concept of "super-human" in the context of artificial intelligence is discussed, notably in games such as Go. Nonetheless, Toral underscores that the simplicity of Go, mainly concerning the number of possible moves at each stage of the game, contrasts sharply with the complexity of human languages, particularly in the domain of translation. Unlike Go, language does not have a clear winner, and there is no single definitive solution. Accordingly, labelling AI systems achieving human-level performance in translation as "human parity" is problematic, as machine translation (MT) operates fundamentally differently from human translators. While AI systems may excel at text cohesion to some extent, achieving discourse external coherence remains beyond their reach due to the requirement of worldly knowledge. This limitation extends to any context-dependent decisions that extend beyond a single sentence.

Hence, human parity and superhuman performance represent controversial issues since, for all the factors just discussed, several criticalities arise when dealing with the evaluation process of translations. Indeed, a closer look at current MT systems demonstrates that they are still far from reaching the aforementioned "human parity". NMT generally considers the

sentence level, despite the efforts of some recent systems trying to include larger contexts. If one considers some specific aspects, the advances in the field are evident. For example, NMT systems, thanks to transformers, managed to assemble different fragments of texts overcoming the problem of ill-formed sentences. Nonetheless, MT still remains quite literal since it is based on knowledge inferred from large collections of parallel data. Moreover, another major issue regards the type of text to translate and the languages involved. In particular, the annual WMT evaluators report that human-like performance is only reachable for some specific language pairs. Overall, it has been argued that MT works better with purely informative texts written in a direct and simple style. On the other hand, when dealing with tasks including different types of texts, terminological issues arise.

5 Productivity of Machine Translated Texts

Concerning the validity and viability of automatic and human translations, one of the main issues to consider is productivity. Technically speaking, with “productivity” we indicate “the ratio of the quantity and quality of units produced to the labour required per unit of time” (O’Brien 2011: 2). Although this definition only looks at the economical aspect of productivity, it conveys the importance of speed of translation in an ever-changing, demanding society looking for the best product obtained in the most limited time. Hence, being productivity one of the major concerns of companies and organisations needing translation services, it is fundamental to better analyse this concept as applied to the translation process. In particular, post-editing productivity also involves the cognitive effort required to achieve the result. In particular, analysing effort in translation means observing how much time and cognitive work is involved during the process. In other words, effort and productivity are inversely related.

Specifically, O’Brien (2011) examines two automatic metrics employed to predict the quality of MT output: i.e. General Text Matcher (GTM) (Turian and Melamed 2003) and Translation Edit Rate (TER) (Snover et al. 2006). GTM metrics assess the similarity between the raw MT output and reference sentence having precision, recall, and their harmonic mean as main criteria. Also, GTM metrics can match adjacent words. According to Turian and Melamed (2003),

the main point of strength of this metric is that it correlates well with the human judgement of adequacy and fluency, two factors of crucial importance in the human evaluation of machine-translated texts (Ma and Cieri 2006).

TER measures the number of edits required to change raw MT output into a reference sentence. The developers of TER tried to achieve the highest correlations with human judgements. The main reason why it was selected for the experiment is that, unlike other metrics, TER does not require a large number of reference sentences to correlate with human judgements. Besides, since the focus of O'Brien was to investigate the effort involved in the translation process, it was selected as this metric records the number of edits necessary to convert raw MT output into a reference sentence. Based on the data obtained from the analysis carried out in the study, the author concludes that there is significant evidence to demonstrate that MT automatic metrics (at least the two under investigation) and actual post-editing productivity do correlate. Hence TER and GTM can be considered reliable metrics that convey post-editing productivity. Nonetheless, the limits of this research are also highlighted, suggesting further investigations that also test them in more detail in terms of the accuracy level of the individual segment. Finally, it would be worth exploring the impact of different language pairs, directions and domains.

6 Towards Standardisation and Customisation of Translation Quality

As regards the translation quality assessment methodologies employed in industry, the development of lists of error types to evaluate translation started to spread in the '90s. One of the most influential models worth reporting is the Localisation Industry Standards Association (LISA) which has continued to be employed in its different adaptations even though it ceased in 2011. LISA is based on a model that includes errors categorised according to three main levels: minor, major, and critical in the opinion of the evaluator. The translation output can be accepted or rejected based on the threshold predefined by the evaluator. That is to say, the status depends on how tolerant or demanding they decide to be. Despite the possibility of customising LISA according to the company's specific requirements, one main drawback of the model is that it does not allow us to have intermediate levels of acceptability of the translation since it can only be either accepted or rejected.

The tendency to objectify translation quality according to quantifiable criteria has led to an urge to standardise the process and, thus, to develop an ISO certification parameter, i.e. the ISO/TS 116699: 2012. It is a guideline standard that serves as guidance concerning best practices for all phases of the translation project. The ISO consists of a framework of 21 parameters classified into five main areas: source content, requirements for the target, production tasks, environment, and relationship. The standards conceive translation quality in these terms: “When both requesters and translation service providers agree on project specifications, the quality of a translation – from workflow and final delivery perspective – can be determined by the degree to which the target content adheres to the predetermined specifications” (ISO/TS 11669: 2012).

However, several scholars disagree with this definition of translation quality. For instance, Koby et al. (2014), to contrast this view, oppose a broad and narrow definition of translation quality. The broad view assumes that there cannot be absolute specifications valid for all translation activities and requirements. On the other hand, the narrow definition focuses on the textual aspect of translation. That is, activities and processes such as summarising, paraphrasing etc. are not considered as part of the translation process. Hence, explicit specifications, according to this view, are often unnecessary because requesters cannot have a clear picture of what a translation project requires. An interesting point to stress is that to evaluate translation quality a proper identification of translation nodes is fundamental and “any effort to measure translation quality is doomed by confusion without an explicit definition of translation quality” (Koby et al. 2014).

An important attempt to develop indicators to achieve a more effective translation quality assessment comes from the Translation Automation User Society (TAUS). Different stakeholders were employed to achieve this goal. Several variables were considered including communicative function, end-user requirements, context, modes of translation (i.e. HT, raw MT output and post-edited MT), content profiling and quality estimation (Castilho et al. 2018). One of the most significant achievements of TAUS is the adoption of the Dynamic Quality Framework (DQF), where quality issues started to be considered before translating.

In the same research context, the EU-funded QTLaunchPad project developed the Multidimensional Quality Metrics (MQM) framework. It provides a flex-

ible way to create and use appropriate metrics for each translation task that can meet both the requester's and users' expected outcomes. The main affordance of the MQM is that it provides a shared metric that can be used for human and machine translation. Based on the identification of over 100 specific translation issues, the metric can be selected by the users depending on the type of project requirements and priorities to support and improve the translation assessment phase. Moreover, once set the specific metric with the selected parameters, it can be stored in a library to be easily reused across similar projects in the future. More specifically, the metric is defined as completing the following tasks:

- Task 1 "Complete specifications": it defines expectations about the translation and serves as the basis for contractual obligations. The 21 parameters included in this task cover all aspects of the translation product, project, and process.
- Task 2 "Select dimensions": MQM dimensions are high-level aspects including fluency, accuracy, verity, design, and internationalisation.
- Task 3 "methods": aimed at minimising human effort, this section includes a basis for the assessment: i.e. analytic, holistic, task-based.
- Task 4 "select issues": for each dimension, a number of related issues must be chosen to measure it following specifications. For example, when assessing fluency, and how linguistically well formed the target or source text is, the issues to be selected may concern spelling, grammar, register, and style. Nonetheless, it must be noticed that the selected issues vary from the genre and text type. For example, style may not be relevant when dealing with technical texts.
- Task 5 "Set issues weights": weights are used to set the relative importance of different issues. For example, terminology may have a different weight than style in certain types of texts.
- Task 6 "Determine thresholds": they can be set per issue or for dimension and are extremely important as they set the criteria of acceptability of the translation output expressed in percentage values.
- Task 7 "implement a workflow": each MQM metric must be implemented in an appropriate workflow with accompanying assessment tools which may include "sanity checks" as well as objective outcomes and decision response: i.e. approved, perform inspection, send back to the translator etc.

7 Conclusion

The paper has addressed the relationship between automatic and human translation, mainly in terms of productivity and quality, taking into account the most influential studies in the field of translation, artificial intelligence and computational linguistics. To investigate this question, a technical overview of the main approaches to MT was provided, focused on how different MT systems work, the paradigms behind their development and the main linguistic criteria included. As regards the controversial issue of translation quality, human parity and super human performance, it can be argued that translation is a complex process involving multiple domains: cognitive, social, cultural, and technological. Hence, finding a unique definition of translation quality that takes into account the multidisciplinary aspect of the process, and capturing its intrinsic complexity is not an easy task. In addition, the discussion on the use of MT, post-editing and automatic metrics shows, essentially, that the post-editing of MT texts has become a part of the translation workflow, raising new important questions in translation research.

What emerges from the analysis of these automatic metrics is that the rise of MT and, consequently, of MT output has contributed to evaluating translation quality a much debated topic in the translation research community. As Lommel et al. (2014) point out, “human and machine translation [...] quality evaluation methods have been fundamentally different in kind, preventing comparison of the two”. If on one hand, it is true that, in the translation industry, quality is mainly related to customer opinion, on the other, it can be noticed a tendency to quantify the assessment process. Indeed, the evaluation models used are mainly error-based, targeted at computing the number of errors detected, classified according to certain standards, and weighted by a reviewer or post-editor. This requires the evaluation system to set the predetermined errors according to logical or hierarchical criteria to be acknowledged and used to provide an effective and objective evaluation of the translation output.

Another crucial point that is worth recalling from the studies reviewed is that post-editing, since its introduction in the translation industry, has obtained a more central position due to the acknowledged benefits in terms

of the productivity of translators. Nonetheless, as human-machine interaction has increased in professional practice, due to the continuing growth in digital content, the aforementioned translation quality has become even more challenging to define, capture, and assess. To operationalise and measure translation quality, different attempts have been made with the aim of achieving evaluation standards based on quantitative criteria (i.e. the ratio of quantity and quality to time). However, the limits of automatic metrics have been discussed pointing to the need to customise the evaluation process, both for human and MT output, according to the specific requirement of the user. It is essential to choose an approach that overcomes dichotomies and is able to join together the two opposite sides of a continuum between the source-oriented concept of fluency and the target-oriented concept of accuracy. Hence, the state of the art suggests that a pragmatic, targeted to end-users method is needed that takes into account the notion of adapting the evaluation system to the purpose of the translation.

To conclude, it can be argued that this overview of methodologies and approaches to translation, post-editing and translation quality assessment, discussing their strengths and weaknesses, is by no means exhaustive. It aims at identifying a number of perceived issues concerning the translation field, in terms of target to achieve, processes, and product evaluation. What arises, particularly for the notion of quality, is that the integration of translation technologies has profoundly changed the relationship between humans and machines, making the boundaries between the two more blurred. Understanding how translation technologies evolve and develop and how the most effective and appropriate evaluation approach can be selected is essential, nowadays, to successfully integrate these technologies in the translation industry. Hence, the ability to adapt to new translation tools and to conceive translation quality with more flexibility and fluidity is crucial considering the impact it may have in terms of effectiveness. Indeed, the translation industry and research “need a way to compare different sorts of translation as objectively as possible, with an emphasis on identifying problems and the metrics adopted to this end should be built on a well-defined foundation including at least clearly stated definitions of translation, quality, and translation quality” (Koby et al. 2014).

References

- Abiola, O.B, Adetunmbi, Adebayo, Oguntimilehin, Abiodun, 2015, “Using hybrid approach for English-to-Yoruba text to text machine translation system (proposed)” *International Journal of Computer Science and Mobile Computing* 4(8), 308-313.
- Akinwale, O.I., Adebayo, Olusola, Adetunmbi, Olumide Olayinka, Obe and A. T. Adeguyi, 2015, “Web-Based English to Yoruba Machine Translation” *International Journal of Language and Linguistics* 3(3), 154-159.
- Bahdanau, Dzmitry, Cho, Kyunghyun, Bengio, Yoshua, 2014, “Neural machine translation by jointly learning to align and translate” *Computation and Language* arXiv preprint arXiv:1409.0473.
- Barrault, Loïc, Bojar, Ondřej, Costa-jussà, Marta R., Federmann, Christian, Fishel, Mark, Graham, Yvette, Haddow, Barry, Huck, Matthias, Koehn, Philippe, Malmasi, Shervin, Monz, Christof, Muller, Matthias, Pal, Santanu, Post, Matt, and Zampieri, Marcos, 2019, “Findings of the 2019 conference on machine translation (WMT19). In Proceedings of the Fourth Conference on Machine Translation” *Florence, Italy, August. Association for Computational Linguistics*, 2: 1-61.
- Belouadi, Jonas and Steffen Eger, 2022, “UScore: An Effective Approach to Fully Unsupervised Evaluation Metrics for Machine Translation.” *ArXiv abs/2202.10062*.
- Bentivogli, Luisa, Bisazza, Arianna, Cettolo, Mauro and Marcello Federico, 2018, “Neural versus phrase-based MT quality: An in-depth analysis on English–German and English–French” *Computer Speech & Language*, 49:52–70.
- Bojar, Ondřej, Federmann, Christian, Fishel, Mark, Graham, Yvette, Haddow, Barry, Koehn, Philipp, and Monz, Christof, 2018, “Findings of the 2018 conference on machine translation (WMT18)” In *Proceedings of the Third Conference on Machine Translation: Shared Task Papers*, pages 272–303, Belgium, Brussels, October. Association for Computational Linguistics.
- Carl, Michael, Dragsted, Barbara, Elming, Jakob, Hardt, Daniel and Arnt Lykke Jakobsen, 2011, “The process of post-editing: a pilot study” in Bernadette Sharp et al. (eds) *Proceedings of the 8th International NLPSC workshop. Special theme: Human-machine interaction in translation*, Copenhagen Studies in Language 41. Frederiksberg, Samfundslitteratur, 131–142.
- Castilho, Sheila, Doherty, Stephen, Gaspari, Federico, Moorkens, Joss, 2018, “Approaches to Human and Machine Translation Quality Assessment” in: Moorkens, J., Castilho, S., Gaspari, F., Doherty, S. (eds) *Translation Quality Assessment. Machine Translation: Technologies and Applications*, vol 1. Springer, Cham.
- Cettolo, Mauro, Niehues, Jan, Stüker, Sebastien, Bentivogli, Luisa and Marcello Federico, 2013, “Report on the 10th IWSLT evaluation campaign” in *Proceedings of the 10th International Workshop on Spoken Language Translation: Evaluation Campaign*, Heidelberg, Germany.

- Edmundson, Harold P. and David G. Hays, 1958, "Research methodology for machine translation" *Mechanical Translation* 5(1), 8-15.
- Fiederer, Rebecca and Sharon O'Brien, 2009, "Quality and machine translation: A realistic objective?" *The Journal of Specialised Translation* 11, 52-74.
- García, Ignacio, 2012, "A brief history of postediting and of research on post-editing" in Anthony Pym and Alexandra Assis Rosa (eds) *New Directions in Translation Studies. Special Issue of Anglo Saxonica* 3(3), 292-310.
- Gaspari, Federico, Almaghout, Hala and Stephen Doherty, 2015, "A survey of machine translation competencies: insights for translation technology educators and practitioners" *Perspectives: Studies in Translatology*, 1-26.
- Hassan, Hany, Aue, Anthony, Chen, Chang, Chowdhary, Vishal, Clark, Jonathan, Federmann, Christian, Huang, Xuedong, Junczys-Dowmunt, Marcin, Lewis, William, Li, Mu, Liu, Shujie, Liu, Tie-Yan, Luo, Renqian, Menezes, Arul, Qin, Tao, Seide, Frank, Tan, Xu, Tian, Fei, Wu, L., Wu, Shuangzhi, Xia, Yingce, Zhang, Dongdong, Zhang, Zhirui, and Ming Zhou, 2018, "Achieving Human Parity on Automatic Chinese to English News Translation" *Microsoft AI & Research* arXiv.
- International Organization for Standardisation 2012 ISO/TS 11669:2012 technical specification: translation projects – general guidance. International Organization for Standardisation, Geneva. Available via: <https://www.iso.org/standard/50687.html>.
- Koby, Geoffrey S., Daryl R. Hague, Arle Lommel and Alan K. Melby, 2014, "Defining translation quality" *Revista Tradumàtica* (12), 413-420.
- Koponen, Maarit, 2016, "Is Machine Translation Post-editing Worth the Effort? A Survey of Research into Post-editing and Effort" *The Journal of Specialised Translation* (25), 131-148.
- Lommel, Arle, Uszkoreit, Hans, Burchardt, Aljoscha, 2014, "Multidimensional Quality Metrics (MQM): a framework for declaring and describing translation quality metrics" *Revista Tradumàtica* 12:455-463.
- Ma X, Cieri, C, 2006 "Corpus support for machine translation at LDC" in *Proceedings of the Fifth International Conference on Language Resources and Evaluation*, Genoa, Italy, 22-28 May, 859- 864.
- Marie, Benjamin, Fujita, Atsushi and Raphael Rubino, 2021, "Scientific credibility of machine translation research: A meta-evaluation of 769 papers" in *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing* (Volume 1: Long Papers), pages 7297-7306, Online. Association for Computational Linguistics.
- Mathur, Nikita, Baldwin, Timothy and Trevor Cohn, 2020, "Tangled up in BLEU: Re-evaluating the Evaluation of Automatic Machine Translation Evaluation Metrics" in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4984-4997, Online. Association for Computational Linguistics.

- O'Brien, Sharon, 2011, "Towards predicting post-editing productivity", *Machine Translation* 25, 197-215.
- Oladosu, John Babalola, Esan, Adebimpe, Adeyanju, Ibrahim, Adegoke, Benjamin, Olaniyan, Olatayo & Omodunbi, Bolaji, 2016, "Approaches to Machine Translation: A Review", *Journal of Engineering and Technology* 1(1), 120-126.
- Ondřej Bojar, Christian Federmann, Mark Fishel, Yvette Graham, Barry Haddow, Matthias Huck, Philipp Koehn, and Christof Monz.
- Papineni, Kishore, Roukos, Salim, Ward, Todd, Zhu, Wei-Jing, 2002, "BLEU: a Method for Automatic Evaluation of Machine Translation" In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics* Stroudsburg PA, USA, 311-318.
- Plitt, Mirko and François Masselot, 2010, "A Productivity Test of Statistical Machine Translation Post-Editing in a Typical Localisation Context", *The Prague Bulletin of Mathematical Linguistics* 93, 7-16.
- Quirk, Chris. & Corston-Oliver, Simon, 2006, "The impact of parse quality on syntactically-informed statistical machine translation" In *Proceedings of Conference on Empirical Methods in Natural Language Processing*, 56-63.
- Snover, Matthew, Dorr, Bonnie, Schwartz, Rich, Micciulla, Linnea, Makhoul, John, 2006, "A study of translation edit rate with targeted human annotation" in *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas*, August 8-12, Cambridge, Massachusetts, USA, 223-231.
- TAUS, 2010, Machine Translation Post-editing Guidelines. <https://www.taus.net/think-tank/best-practices/postedit-best-practices/machine-translation-post-editing-guidelines>.
- Thang, Luong, Hieu, Pham, and Christopher D. Manning, 2015, "Effective Approaches to Attention-based Neural Machine Translation" in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 1412-1421, Lisbon, Portugal. Association for Computational Linguistics.
- Toral, Antonio, 2020, "Reassessing claims of human parity and super-human performance in machine translation at WMT 2019" in *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, pages 185-194, Lisboa, Portugal, November. European Association for Machine Translation.
- Toral, Antonio, Castilho, Sheila, Hu, Ke, and Andy Way, 2018, "Attaining the unattainable? Reassessing claims of human parity in neural machine translation" in *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 113-123, Brussels, Belgium, October. Association for Computational Linguistics.
- Turian, Joseph, Shen, Luke, Melamed, I. Dan, 2003, "Evaluation of machine translation and its evaluation" in *Proceedings of the MT Summit IX*, New Orleans, USA, 23-27 September 2003, 386-393.

LUCIA DI PACE, ROSSELLA PANNAIN*

Strategie linguistiche nella campagna elettorale delle politiche 2022. Una focalizzazione sull'impiego dei deittici di persona

Abstract

Il lavoro propone un'analisi dei manifesti prodotti da tre fra le maggiori forze politiche italiane in occasione della campagna elettorale per le politiche del 25 settembre 2022. L'obiettivo primario è stato quello di evidenziare come i tre schieramenti abbiano sfruttato, in maniere divergenti, i meccanismi di riferimento ai partecipanti della dimensione comunicativa, creando effetti tanto di avvicinamento, coincidenti in alcuni casi con la sovrapposizione inclusiva, quanto di distanziamento tra locutore e interlocutore. Un ulteriore obiettivo è stato verificare se nei testi dei manifesti elettorali si ritrovino aspetti caratterizzanti del discorso politico contemporaneo e in quale misura essi possano essere associati alla categoria dello "slogan" politico come di quello pubblicitario.

Parole chiave: deissi di persona, inclusivo/esclusivo, linguaggio della politica/politico, slogan, manifesto elettorale

The paper proposes an analysis of the election posters produced by three among the major Italian political forces, in the campaign for the general elections of September 25, 2022. The primary object of the study was to investigate how the three parties exploited the mechanisms of reference to participants in the communicative situation, resulting in either effects of closeness, up to the inclusive identification, or effects of distancing between locutor and interlocutor. A further aim was to determine if the texts of the election posters manifest specific characteristics of contemporary political discourse, and to what extent they may be associated to the category of "slogans", political as well as commercial.

Keywords: personal deixis, inclusive/exclusive, language of politics, slogan, election poster

* Lucia di Pace, Università di Napoli L'Orientale, ldipace@unior.it; Rossella Pannain, Università di Napoli L'Orientale, rpannain@unior.it.

1. Introduzione

Il profondo cambiamento verificatosi nel linguaggio della politica, dall'inizio della Seconda Repubblica fino ai giorni nostri, è ormai stato analizzato da più parti, tanto con lavori che si dedicano a cogliere aspetti generali e tendenze condivise quanto con studi volti ad analizzare gli stili discorsivi di alcuni esponenti politici di rilievo (da Berlusconi a Bossi, da Renzi a Grillo)¹. In particolare, è stato messo in evidenza come il discorso politico sia passato a una dimensione che si può definire di "mediatizzazione", dal momento che viene veicolato, oltre che nei canali più tradizionali e istituzionali, in maniera ormai preponderante attraverso la TV e il web, pur continuando ad utilizzare forme preesistenti come manifesti, volantini, e altre forme di propaganda itinerante come il pullman elettorale. Questo aspetto è determinante per spiegare sia il mutamento in termini strettamente linguistici, che Antonelli (2020) definisce come passaggio dal "paradigma della superiorità" al "paradigma del rispecchiamento", vale a dire il passaggio da un linguaggio astruso e inaccessibile (il cosiddetto "politichese") ad uno più semplice nel quale il cittadino possa rispecchiarsi, sia l'emergere e il consolidarsi di caratteristiche nuove della comunicazione politica. Queste ultime sono colte in lavori di riferimento ormai imprescindibili (Gualdo & Dell'Anna 2004; Dell'Anna & Lala 2004): la spettacolarizzazione e le operazioni di marketing politico, legate alla dimensione mediatica (v., ad es., Calabrese 1998), la personalizzazione e il leaderismo, la tendenza ad una comunicazione povera, l'impiego della dinamica "vecchio"/"nuovo". Paola Desideri rimarca come il nuovo stile comunicativo manifesti "una potenza mediatica molto simile a quella pubblicitaria" (Desideri 2011), rafforzando il carattere della competizione accesa che si instaura tra i rappresentanti politici dei diversi schieramenti. In realtà, si può affermare che la politica non semplicemente si ispiri al codice della pubblicità, ma ne faccia uso in senso vero e proprio: tutti i partiti si affidano a professionisti del marketing e della comunicazione d'impresa, realizzando la cosiddetta "promocrazia" (Bruno 1996); non a caso, questa tendenza della comunicazione politica più recente è stata sarcasticamente etichettata come

¹ La mole di lavori è tale che, in questa sede, non si può pensare di farne nemmeno un rapido resoconto.

“SpotPolitik” (Cosenza 2012)². Inoltre, l’avvicinamento alla dimensione pubblicitaria, che prevede l’aperta contrapposizione con il concorrente - ormai consentita a partire dal 2000 con la legalizzazione della pubblicità comparativa - consente di rilevare una sorta di rafforzamento della strategia discorsiva ben nota in ambito politico, l’*argumentum ad personam* (De Santis 2019: 365)³.

Se queste sono tra le caratteristiche colte nel linguaggio della politica dagli anni ’90 ad oggi, non sorprende che, in particolare nelle campagne elettorali, esse possano trovare un’ulteriore conferma. Si noti, peraltro, che il termine stesso “campagna” si addice tanto alla dimensione politica quanto a quella pubblicitaria, evocando lo scenario di una competizione/battaglia: la campagna elettorale è una campagna pubblicitaria.

Un primo obiettivo di questo lavoro, incentrato sull’analisi linguistica dei manifesti della campagna elettorale per le politiche del 25 settembre 2022, è mostrare come gli slogan, o semplicemente i testi, usati nei manifesti elettorali, riproducano gli stessi caratteri sopra ricordati, con una forza e incisività anche maggiori. In effetti, come nota Sergio (2008: 27), i testi presenti sui manifesti elettorali possono essere considerati come condensati di discorsi, o come discorsi “in pillole”.

Allo stesso tempo, va sottolineato come gli slogan, forma tipica della comunicazione in fase di campagna elettorale, rappresentino una tipologia testuale particolare, che ha caratteri propri e, per certi aspetti, diversi dai veri e propri discorsi politici. Lo slogan politico, come quello commerciale o quello urlato dai tifosi o esibito in forma scritta sui cartelloni negli stadi, è una testualità estremamente breve, non presenta alcun tratto di ricercatezza lessicale, ha carattere formulaico, sfrutta meccanismi eufonici funzionali alla facile memorizzazione (Viviani 2011; De Santis 2019). In realtà, se si tiene conto di una opportuna distinzione, all’interno degli slogan politici, tra messaggi di propaganda e messaggi di contestazione di piazza, si coglie come questi ultimi in particolare aderiscano al tipo testuale dello slogan, dal momento che in genere vengono enunciati o addirittura cantati in una dimen-

² L’autrice denuncia il fatto che la strategia comunicativa dei nuovi leader si riduca alla scelta di uno “slogan generico, due colori per il logo e qualche foto per le affissioni” (*Premessa*, p. 1).

³ Cristiana De Santis fa riferimento a questa motivazione per illustrare in particolare la contrapposizione con il proprio avversario politico sui manifesti elettorali.

sione corale, manifestando una maggiore dipendenza da meccanismi di tipo fonico. Al contrario, i testi presenti nei manifesti elettorali possono essere considerati per certi aspetti lontani dagli slogan in senso stretto, puntando soprattutto alla trasmissione di un contenuto – in molti casi solo evocato –, pur conservando il carattere della brevità, tipico di questo genere. Di fatto, come si osserverà, le testualità presenti nei recenti manifesti elettorali mostrano, rispetto al passato, un sempre minor grado di raffinatezza testuale e, soprattutto, fanno poco ricorso all’impiego di espedienti retorici canonici, riducendo il peso della funzione poetica, che, con effetti stilistici particolari o di “straniamento”, dovrebbe caratterizzare questa tipologia testuale.

In quest’ottica, l’obiettivo principale di questo studio sarà di mostrare come i testi dei manifesti elettorali mettano in secondo piano la strutturazione vera e propria del messaggio – venendosi a configurare in questo senso come “slogan deboli” – per concentrarsi invece sui meccanismi di riferimento ai partecipanti della dimensione comunicativa propria del discorso politico elettorale: il mittente (il leader politico, il partito, lo schieramento politico) e il destinatario (l’elettore, il sostenitore), contrapposti alla realtà designata, che include l’altra parte politica, l’avversario, l’altro. Come noto (v., in particolare, Desideri 1984: 23-24 e 1999: 400), il discorso politico si caratterizza nell’essere enunciazione rispetto ad un enunciato, chiamando in causa fondamentali meccanismi di deissi, tanto di persona quanto di spazio e tempo. I deittici consentono i meccanismi di *embrayage* e *débrayage* attraverso i quali il mittente, il soggetto dell’enunciazione, àncora il proprio messaggio alla dimensione dell’enunciazione: l’uso dell’“io”, come del “qui”, dell’“ora”, consente una presa di contatto diretta con il proprio uditorio, mentre l’uso della III persona crea una maggiore distanza. Come si osserverà, la strutturazione linguistica degli slogan creati durante la recente campagna elettorale si costruisce giocando sull’alternanza di processi di avvicinamento e allontanamento tra locutore e interlocutore, con focalizzazioni sui diversi ruoli e con meccanismi di rispecchiamento e coincidenza tra i diversi attori della situazione comunicativa.

I diversi punti di vista da cui veicolare il messaggio, quello del mittente o del destinatario, si contrappongono a prospettive in cui diventa centrale il contenuto del messaggio, presentato in uno scenario di oggettività. A tali scelte si sovrappongono quindi quelle che, in termini jakobsoniani, vedono prevalere una funzione emotiva o espressiva (mittente), conativa (destina-

tario) o referenziale (contesto, stato dei fatti), laddove, come anticipato, la funzione poetica è decisamente marginalizzata.

Questo lavoro si propone, attraverso un approccio qualitativo piuttosto che quantitativo, di individuare ed analizzare tali tendenze, riscontrabili nei manifesti elettorali delle recenti elezioni politiche del 25 settembre, basandosi su un campione di testualità che si ritiene rappresentativo del ventaglio delle scelte comunicative operate dalle forze politiche. L'analisi non è stata condotta in maniera sistematica in relazione a tutte le componenti politiche presenti nello scontro elettorale, ma si è focalizzata su tre forze politiche maggiori che, di fatto, si sono rivelate fortemente divergenti nelle opzioni comunicative effettuate. Il materiale utilizzato è stato reperito sul web, attingendo a siti dei partiti, e consiste per la maggior parte di manifesti che sono circolati pubblicamente sotto forma di cartelloni; tuttavia, si è fatto ricorso anche a manifesti pubblicati su canali "social" come Facebook e Twitter, tanto dei partiti quanto dei leader politici, in considerazione del fatto che questi rappresentano ormai luoghi canonici di comunicazione politica (v., tra gli altri, Amenta 2011 e Spina 2012). Infine, alcuni dei manifesti politici sono stati tratti da archivi dedicati, come la banca dati *manifestipolitici.it* (<https://www.manifestipolitici.it/>) e *l'Archivio degli Spot Politici e dei manifesti* (<http://www.archivispotpolitici.it/>)⁴.

2. La categoria di persona

È utile ricordare che la deissi di persona si manifesta primariamente nei pronomi personali, considerati deittici per eccellenza (vedi ad es. Uspenskij 2008: 112), ma anche nella categoria grammaticale di persona del verbo; non a caso Benveniste, come ci ricorda Manetti (2015: 24), inizia la sua riflessione sulla "persona" in un lavoro dedicato all'indicazione di questa categoria nel verbo (1946), per passare, dieci anni dopo (1956), a rivolgere la propria attenzione

⁴ Alcuni dei manifesti verranno riprodotti nella loro interezza, mentre per altri sono riportati i soli contenuti verbali, riproducendo l'opzione grafica maiuscolo/minuscolo e i segni di punteggiatura, mentre l'andata a capo è segnalata da un trattino. La presenza del simbolo del partito viene indicata tra parentesi quadre, ad es.: [simbolo Lega/SALVINI-PREMIER]. Per i manifesti presentati anche in forma di immagine, qualora siano stati tratti da profili social, o dai due archivi di manifesti indicati nell'Introduzione, ciò viene esplicitato nel testo oppure in nota.

alla natura dei pronomi, dando origine alla teoria dell'enunciazione (1956).⁵ La categoria di persona emerge anche nei possessivi: la funzione di questi ultimi, infatti, non è tanto quella di esprimere il possesso, quanto piuttosto di indicare le relazioni tra le tre persone in riferimento a ciò che viene enunciato (Simone 2013: 173). Inoltre, i ruoli di mittente e destinatario emergono necessariamente in determinate categorie modali del verbo, come l'imperativo.

2.1 I pronomi personali

Come è stato notato da Benveniste (1966a, 1966b), e successivamente ribadito da innumerevoli studiosi, anche contemporanei come Bhat (2004) e Daniel (2020), i pronomi personali codificano il riferimento a un determinato ruolo nell'atto comunicativo: il parlante, il ricevente, o una combinazione di questi. Il riferimento a nessuno di questi due ruoli è veicolato dalla terza persona, che, proprio nel non essere partecipante all'interazione comunicativa ("non-locutore", Daniel 2020), viene definita negativamente come "non-persona" (Benveniste 1966b: 228, Lyons 1977: 638), o "persona assente" (etichetta presente nella tradizione grammaticale araba alla quale Benveniste si ispira). In effetti la III persona è concepita come un'entità separata dalle altre due: "è la sola attraverso la quale si predica una cosa verbalmente" (Benveniste 1966b: 230). La natura profondamente diversa tra i pronomi di I e II, da una parte, e quello di III, dall'altra, è colta anche da Bhat (2004: 2-3), che sottolinea la problematicità insita nella definizione della categoria generale del pronome, "ciò che sta per un nome", e osserva come essa sia del tutto inadeguata per i pronomi personali di I e II persona. Egli propone la distinzione, su base funzionale, tra i "pronomi personali" (solo quelli di I e II persona) e le "proforme", categoria che include tutti gli altri, compresa la III persona dei personali di molte lingue. Solo i primi hanno come funzione primaria quella di indicare i due ruoli di parlante e ricevente, senza per altro fornire alcuna base per l'identificazione degli effettivi individui che svolgono quei ruoli. Non a caso, nella maggioranza delle lingue del mondo, il genere non si manifesta nella I e II persona, ma solo nella III. Infatti, se una lingua ha distinzione di genere per le persone diverse dalla terza, allora la ha anche per la terza, come emerge dalla rie-

⁵ I due scritti di Benveniste saranno menzionati in questo lavoro facendo riferimento alla data della loro successiva pubblicazione nei *Problèmes de linguistique générale* (1966).

laborazione dell'Universale 44 di Greenberg da parte di Siewierska (2013). Questo fenomeno può avere una base funzionale, dal momento che la III, diversamente dalle prime due, è potenzialmente ambigua nella sua referenza (Corbett 1991: 131). Conferme dello status particolare della III persona sono rappresentate dal fatto, anche questo già notato da Benveniste (1966b: 228-229), che nella maggior parte delle lingue flessive la forma verbale di III singolare si presenta con il solo tema, senza affisso di persona e dalla circostanza che i verbi cosiddetti "impersonali" si manifestano in realtà proprio alla III persona singolare.

Quando si passa ad analizzare le forme plurali dei pronomi personali, emergono altri aspetti del tutto particolari: come già evidenziava Benveniste (1966: 233), la natura stessa dei pronomi personali impedisce la possibilità di una pluralizzazione simile a quella che si ha per i nomi. Di fatto, la maggior parte delle lingue del mondo – tra cui l'italiano – esprime la persona e il numero in modo congiunto, cioè in forme lessicali dedicate, piuttosto che attraverso morfemi separati (tra gli altri, Huddleston & Pullum 2002: 1465 e Daniel 2020: 31). Delle 216 lingue che costituiscono il campione analizzato da Daniel nel WALSH (2013), solo 40 non seguono questo modello, come ad es. il cinese, che usa affissi aggiunti alla forma singolare, o lo yukaghiro, che fa ricorso a variazione apofonica (ad es. *met* 'io' e *mit* 'noi'; *tet* 'tu' e *tit* 'voi'). Questo tratto di divaricazione tra nomi e pronomi personali sulla base del criterio della pluralizzazione, come osservato da Jespersen (1924: 192) e Lyons (1968: 277), rende la terminologia grammaticale tradizionale fuorviante (Baht 2004: 92). Da questa prospettiva, appare più appropriata la scelta di autori come Serianni (1988) che preferiscono adottare una terminologia metalinguistica che dispone le forme pronominali in una sequenza ordinale, evitando lo schema singolare/plurale, così che il "noi", ad esempio, non è più la I persona plurale, ma la "quarta persona".

La I persona plurale in particolare si configura come una categoria che sfugge al semplice gioco delle corrispondenze sulla base del tratto "numero", tanto da essere stata addirittura definita come "monstrum" (La Fauci 2016: 392). In effetti, come già più volte rilevato nella letteratura degli studi su grammatica e pragmatica (Uspenskij 2008: 119-120, ad esempio, richiama autori come Bopp, Jespersen, Vinogradov), la prima persona plurale non può essere considerata un plurale della I singolare "io", dal momento che l'io, nell'essere 'la persona che parla', non è pluralizzabile o ripetibile. Di conse-

guenza, la referenza di *noi* non sarà ‘io + io (+ io)...’, ma o ‘io + tu’ o ‘io + voi’, oppure, in modo non inclusivo, ‘io + lei/lui’ o ancora ‘io + loro’, quindi il parlante e uno o più non parlanti (Baht 2004: 92). Jespersen (1924: 191) considera il pronome ‘noi’ come un plurale di approssimazione in cui diversi individui sono compresi nella stessa designazione, pur non appartenendo esattamente allo stesso tipo. D’altra parte, anche il plurale della seconda persona è caratterizzato da tratti di indefinitezza: può indicare, in alcuni contesti, più riceventi (configurandosi in tal senso come un plurale del “tu”), ma generalmente indica il ricevente e una o più altre persone che non lo sono (Baht 2004: 92).

Un ulteriore modo in cui i pronomi, e in particolare quello di I persona, si differenziano da altri nominali è la possibilità di esprimere i valori di inclusivo/esclusivo (Bhat 2004: 92; Cysouw 2013). Nel momento in cui si utilizza il pronome personale *noi* con un’accezione inclusiva si postula la condivisione di un’esperienza tra parlante e interlocutore/i, che in tal modo diventa realtà oggettiva.

In una lingua come l’italiano, che non opera la distinzione inclusivo/esclusivo (codificata grammaticalmente in altre lingue), tranne che tramite le forme analitiche, poco usate e appartenenti a un registro non standard, *noi altri/voi altri*⁶, la I persona plurale, definita da Daniel (2020) come “noi-neutro”, è inerentemente ambigua quanto a riferimento, giacché può essere intesa/interpretata come includente oppure escludente il ricevente. In questo senso il *noi*, in italiano, ha una capacità di variare il proprio riferimento più ampia rispetto agli altri pronomi, già globalmente definiti *shifters* da Jakobson (1957) e, con riferimento specifico alla I e II persona, da Bhat (2004: 2).

Dei due possibili riferimenti di I persona plurale, quello inclusivo è stato oggetto di particolare attenzione da parte degli studiosi, configurandosi come una categoria cui si associa una specifica valenza pragmatica. Anche nelle descrizioni grammaticali dell’italiano ci si sofferma su usi particolari del pronome di I persona plurale, fin qui definiti “inclusivi”: il “plurale sociativo”, o di “coinvolgimento” (Renzi 2001: 354), il plurale “didattico”, quello “narrativo” (Serianni 1988: 210-11), che hanno in comune l’obiettivo di coinvolgere o convincere il ricevente, rendendolo parte della categoria del locutore. I mittenti tipici di testualità nelle quali si riscontra questo uso della

⁶ <https://www.treccani.it/vocabolario/noialtri/>.

I persona plurale si trovano in una relazione asimmetrica rispetto al destinatario; tra questi, gli insegnanti, gli educatori, i sacerdoti, le figure istituzionali e i politici. Come si mostrerà, nel discorso politico, il plurale narrativo, in particolare, è funzionale alla creazione di un rapporto empatico con gli interlocutori, portandoli a condividere sentimenti, giudizi, valutazioni.

Ma le possibilità referenziali del “noi” non terminano qui: la polarità inclusivo/esclusivo, canonicamente individuata in relazione al destinatario (“io” + “tu”/ “voi” opposto a “io”+ “lui”/ “loro”) si può cogliere anche in riferimento al mittente; si può quindi sostenere che esista un “noi” inclusivo del parlante e un “noi” esclusivo del parlante. In realtà, l’esclusione dell’“io” si riconosce in usi assolutamente marginali che, ancora una volta, si manifestano in situazioni di comunicazione asimmetrica, nelle quali si parla ad un bambino, ad una persona malata o anziana. Sono casi in cui l’“io” si cancella nel “noi”, si auto-esclude, in quanto si identifica con un “io” diverso a cui dà voce: si pensi ai casi in cui i genitori producono enunciati per i quali il mittente inteso è il bambino (*abbiamo mangiato tutto*), similmente a quanto accade nell’interazione medico-paziente (*adesso prendiamo un’altra medicina*). D’altra parte, sulle pratiche di divaricazione tra parlante e mittente, e ascoltatore e destinatario, la pragmatica ha già da tempo puntato l’attenzione (si veda ad es., il modello SPEAKING di Hymes 1974).

Lo statuto del tutto particolare del pronome “noi”, con il suo alto tasso di versatilità, spiega perché sia stato particolarmente impiegato nel discorso politico, ricevendo una parallela attenzione da parte degli studiosi. Di fatto, all’impiego del “noi” in politica sono stati dedicati numerosi lavori: per quanto concerne l’ambito italiano si possono segnalare, a titolo esemplificativo, il contributo di Santulli, che parla di un “uso cruciale della prima persona politica” (2005: 119), di La Fauci, che lo mette ad esponente del titolo di un suo saggio *Noi, persona politica* (2016), di De Santis, che definisce “il pronome personale noi, pronome politico per eccellenza” (2016: 329).

2.1 Il modo imperativo

Negli slogan elettorali che andremo ad esaminare, come è prevedibile in uno scenario in cui ci si appella direttamente al proprio interlocutore, emergono frequentemente forme di imperativo, tanto alla seconda persona singolare

(più diretta, con coinvolgimento individuale del singolo elettore) quanto a quella plurale (con riferimento ad un meno distinto corpo elettorale). Tuttavia, si riscontrano anche usi con la prima persona plurale, le cui implicazioni pragmatiche appaiono particolarmente interessanti nella prospettiva del rapporto che il mittente va a instaurare con il destinatario.

Come noto, in italiano l'imperativo ha forme proprie soltanto per la seconda persona singolare e plurale, mentre nelle altre persone, vale a dire per la III persona singolare e plurale e per la I plurale, è supplito da forme del congiuntivo in funzione esortativa (si veda ad es. Serianni 1988: 401-2). Ancora una volta, la forma relativa alla I persona plurale manifesta una peculiarità rispetto alle altre: la forma della I persona plurale del congiuntivo presente, di tutti i verbi italiani, è uguale alla forma della I persona dell'indicativo presente (*cantiamo, temiamo, dormiamo*), in virtù di un processo diacronico che, dal latino, ha portato alla sostituzione delle forme dell'indicativo con quelle del congiuntivo nella loro interezza (Vincent 1980: 389-90). Si può sostenere che la forma del congiuntivo ha quindi occupato la cella tanto dell'imperativo quanto dell'indicativo presente, determinando probabilmente un'associazione tra le due forme: l'imperativo di I plurale è avvertito come caratterizzato da un minor tasso di "imperatività".

La coincidenza del significante nelle forme di I persona plurale dell'imperativo e del presente indicativo, può, in determinati contesti, determinare un'ambiguità di interpretazioni, veicolando alternativamente un valore inclusivo, quello dell'imperativo, o neutro e potenzialmente esclusivo, quello dell'indicativo. Un'ambiguità che può essere sfruttata a fini pragmatici, in particolare in messaggi la cui ricezione avviene a distanza dalla produzione, brevi e spesso ellittici, come gli slogan pubblicitari.

Da una prospettiva cross-linguistica si può notare (Dobrushina & Goussev 2005: 179-180) come proprio l'imperativo, tra i diversi modi verbali, tenda ad implicare la nozione di inclusione: la categoria verbale dell'imperativo è per sua natura orientata verso il destinatario ed è quindi difficilmente compatibile con una semantica esclusiva. Il tratto dell'inclusività emerge, in particolare, nella I forma plurale: essa veicola di fatto l'invito rivolto a un destinatario a compiere un'azione insieme al mittente, tant'è che in molte lingue esistono solo forme inclusive. Del resto, l'invito ad un'azione comune è tra le attività linguistiche più comuni e salienti, in quanto cruciale ai fini della cooperazio-

ne tra gli esseri umani: in base alla loro ricognizione su larga scala, Dobrushina & Goussev constatano che non si riscontrano lingue che non abbiano una qualche forma per realizzarla. Infatti, l'imperativo di I persona rappresenta la manifestazione più diffusa dell'inclusività nelle lingue del mondo.

È interessante notare come la contrapposizione tra forme inclusive ed esclusive dell'imperativo si riscontri in alcuni modi dire, tra cui il noto *armiamoci e partite!*. Nell'interpretazione di questa frase si tende a mettere in evidenza la polarizzazione tra le due forme verbali, con riferimento alla semantica dei verbi; tuttavia, è evidente come il contrasto sia realizzato proprio dalla marcatura di persona presente nei verbi (un "noi" inclusivo contrapposto ad un "voi").

3. I leader politici nei manifesti della campagna elettorale 2022

Nei paragrafi che seguono ci soffermeremo sull'analisi degli slogan delle ultime elezioni politiche per mostrare quale ruolo vi giochino i meccanismi di riferimento personale: si vedrà come i messaggi elettorali utilizzati da alcuni leader siano orientati su una singola opzione, come nel caso di Matteo Salvini, mentre altri, in particolare quelli di Giorgia Meloni, impieghino un ampio ventaglio delle opzioni disponibili, realizzando giochi di contrapposizione e/o identificazione tra i vari ruoli.

La leader di Fratelli d'Italia, come si mostrerà, è incline ad utilizzare in particolare forme di riferimento plurale con l'obiettivo di coinvolgere l'elettore. Non a caso, farà uso in particolare del "noi" che, come è stato già evidenziato ad esempio nella comunicazione politica di Renzi (si vedano De Santis 2016 e De Santis & Simonini 2017), mostra tutta la sua polivalenza nell'indicare tanto un "io maiestatico", quanto un "noi inclusivo e/o esclusivo".

3.1. Salvini e il dominio dell'io

La scelta di Salvini di basare gli slogan politici elettorali sull'impiego della I persona, attraverso la forma verbale *Credo*, si iscrive nella tendenza alla personalizzazione e auto-narrazione, tipica del discorso politico degli ultimi anni, riscontrabile in particolare in politici della destra (si vedano ad esempio Desideri 2011 e Dell'Anna 2017, entrambe con riferimento a Berlusconi). Nella comunicazione politica di Berlusconi, come è stato sottolineato da più parti, si riconosce un tipo di discorso imperniato sul sé, una sorta di discorso auto-riflessivo, che

Salvini riproduce in pieno. È interessante notare come Berlusconi, e non Bossi, possa essere considerato l'antecedente di Salvini: d'altra parte è con Berlusconi che si è passati da un modello di comunicazione partito-centrico ad uno candidato-centrico, seguendo un modello statunitense (Campus 2004).

Un'altra conferma della tendenza alla personalizzazione è rappresentata dal ritorno, a partire dagli anni '80, dell'impiego dell'immagine del leader nei manifesti politici (Sergio 2008); molto probabilmente, come suggerisce De Santis (2019: 363), l'uso dell'immagine del politico, come strategia di propaganda elettorale, era stata per lungo tempo evitata a causa di una sorta di "complesso del dittatore". In realtà, nei manifesti elettorali delle ultime elezioni politiche, la gran parte degli schieramenti ha riproposto l'immagine del proprio leader, dalla Lega, a Forza Italia, al Movimento 5 Stelle, a Fratelli d'Italia e, in seconda battuta, anche il Partito Democratico, con pochissime eccezioni in controtendenza. La Lega propone pertanto sistematicamente una foto del leader Matteo Salvini in bella evidenza, con pugno stretto a simboleggiare la forza, e anche la vittoria, rafforzata proprio da questa incisiva affermazione linguistica del sé (Fig. 1). Inoltre, si può sottolineare, come ulteriore indice dell'elevato tasso di personalizzazione, che il simbolo del partito include il nome di Salvini proposto come *premier*, in modo solo parzialmente sovrapponibile alla scelta di FdI⁷, laddove, il simbolo del PD non riporta alcun riferimento a persone.



Fig. 1

Fig. 2⁸

⁷ Si veda più avanti, paragrafo 3.3.

⁸ Manifesto pubblicato sull'account Twitter Lega-Salvini Premier il 18.08.2022.

Nella dinamica dei giochi di ruolo tra mittente e destinatario è evidente la salienza e dominanza dell'“io”, rispecchiata nel corpo ben maggiore dei caratteri che compongono la parola *Credo*, rispetto all'emergere degli altri attori, “tu” e “noi”, presenti nella testualità in fondo al manifesto: *Unisciti a noi*. In quest'ultima formulazione si riconosce la funzione d'appello realizzata con la forma imperativa più canonica, alla II persona singolare, resa ancor più incisiva dalla presenza del clitico *ti*, dove il *noi* si configura con ogni evidenza come un 'noi' esclusivo ('noi della Lega'). Si noti che questo schema è stato utilizzato spesso nelle campagne elettorali di tesseramento ai partiti. A titolo esemplificativo, si ricordano gli slogan presenti nei manifesti di Forza Italia delle campagne di tesseramento 1998 e 2004, in cui ricompare la frase *Unisciti a noi*:



Fig. 3

Fig. 4⁹

Dal punto di vista dei contenuti veicolati nello slogan salviniano, così breve e essenziale, si può notare la retorica nel collocare come oggetto del suo “credo”, non un ideale, un proposito, ma gli italiani stessi, venendo a realizzare così una sorta di patto tra sé e gli elettori. L'accento su “gli italiani” è ulteriormente marcato nel manifesto in Fig. 2, nella cui parte inferiore compare la formula *Prima gli italiani*, rafforzata dai colori dell'Italia, che, presumibil-

⁹ Questi due manifesti sono stati tratti dalla banca dati manifestipolitici.it, <https://www.manifestipolitici.it/SebinaOpacGramsci/.do>.

mente, li pone in contrasto con i “non italiani” (su questa implicazione e sulla catena di significati associata a questa formula, già usata da Salvini, si veda De Santis 2019: 368).

Peraltro, si ricorda che lo stesso slogan, *credo*, ancora più enfatizzato con ridondante presenza del pronome, era già apparso per le elezioni politiche del 2008 in un manifesto elettorale della destra (Fig. 5), in cui in luogo de “gli italiani” troviamo il “popolo italiano”

Fig. 5¹⁰Fig. 6¹¹

L'accento sugli italiani, o l'Italia (altro tratto caratterizzante della politica di destra, evidentemente condiviso con Forza Italia e Fratelli d'Italia, persino nel nome del partito) è presente in un manifesto apparso sulla pagina FB di Salvini, in cui apparentemente il centro del messaggio non è Salvini stesso, ma gli elettori, denotati come “italiani”: *il 25 settembre votano gli italiani!*. Eppure è sempre lui, per quanto di spalle, il tema della predicazione, ripreso nei titoli degli articoli riportati in alto (Fig. 6).

Naturalmente, il “Credo” salviniano si riempie di contenuti politici in altri manifesti in cui compaiono ancora *gli italiani*, *i giovani italiani*, o *l'Italia sicura*, o ancora *l'Italia pulita*, e *nessun italiano* (si fa notare che tutti questi rife-

¹⁰ Manifesto tratto dall'Archivio degli Spot Politici e dei manifesti, <http://www.archivispotpolitici.it/>.

¹¹ Manifesto pubblicato sul profilo Facebook di Salvini il 16.09.2022.

rimenti compaiono in modo enfatico, con caratteri maiuscoli) coniugati con i temi classici della politica leghista:



Fig. 7



Fig. 8

STOP SBARCHI - Credo NELL'ITALIA SICURA - Lo abbiamo fatto e lo rifaremo - [simbolo Lega/SALVINI-PREMIER] - #25settebrevotoLega - Uniscitiano: legasonline.it/iocicredo

PACE FISCALE - FLAT TAX 15% - Credo NEGLI ITALIANI - Fermiamo insieme la sinistra delle tasse! - [simbolo Lega/SALVINI-PREMIER] - #25settebrevotoLega - Uniscitiano: legasonline.it/iocicredo

STOP FORNERO! - SÌ QUOTA 41 - Credo - in pensioni giuste e spazio ai - GIOVANI ITALIANI - simbolo Lega/SALVINI-PREMIER] - #25settebrevotoLega - Uniscitiano: legasonline.it/iocicredo

INDIPENDENZA - ENERGETICA E - NUCLEARE SICURO - Credo - NELL'ITALIA PULITA - che coniuga - crescita e ambiente - [simbolo Lega/SALVINI-PREMIER] - Uniscitiano: legasonline.it/iocicredo #25settebrevotoLega

IVA ZERO - SU PANE, PASTA, RISO, - LATTE, FRUTTA E VERDURA - Credo - CHE NESSUN ITALIANO - vada lasciato indietro - [simbolo Lega/SALVINI-PREMIER] - #25settebrevotoLega - Uniscitiano: legasonline.it/iocicredo.

In questi manifesti, che hanno avuto grandissima circolazione sotto forma di cartelloni, emergono, oltre all'“io” salviniano, diverse forme del “noi”: nel manifesto sugli sbarchi (Fig. 7) si può osservare un “noi” con caratterizzazio-

ne esclusiva, in quanto parte politica: *lo abbiamo fatto e lo rifaremo*; si può anche notare come la forza di ancoraggio del testo alla situazione comunicativa si realizzi attraverso il ricorso alla deissi temporale, oltre che personale, delle forme verbali. Inoltre in questo testo, come in quello sulle pensioni, si può notare la forza dell'argomento "contro", realizzato con un linguaggio telegrafico e quasi violento nella mancanza di sintassi: *Stop sbarchi; Stop Fornero!*

Nel manifesto in Fig. 8 si manifesta l'altra dimensione del "noi", quella inclusiva, *Fermiamo insieme la sinistra delle tasse!*, realizzata attraverso l'imperativo esortativo (corroborato dal punto esclamativo) che, come argomentato sopra (2.1), la implica come tratto intrinseco nelle forme di I plurale, e rafforzata dalla presenza dell'avverbio *insieme*. Anche in questo manifesto emerge un'occorrenza della abusata strategia retorica tipica del discorso politico: l'*argumentum ad personam*, per quanto rappresentato nello specifico da un'intera parte politica (*la sinistra delle tasse*). Il "noi" si configura quindi come inclusivo in quanto somma della parte politica e dell'elettorato (mittente + destinatario), mentre esclude la terza persona, l'altro con cui si contrappone.

L'atteggiamento inclusivo e accogliente può essere veicolato con mezzi linguistici meno sottili, come nell'esplicito *Credo che nessun italiano vada lasciato indietro*: in questo manifesto non si può non notare l'abbassamento del livello della comunicazione politica (*Iva zero su pane, pasta, riso, latte, frutta e verdura*) che, per quanto necessariamente sintetica nei manifesti elettorali, può almeno alludere ad argomentazioni politiche, e che qui sembra invece evocare solo la lista della spesa, raggiungendo forse il punto estremo del cosiddetto paradigma del "rispecchiamento" (Antonelli 2000), realizzando appieno il "gentese" non solo nello stile, ma anche nei contenuti (sul "gentese" contrapposto al "politichese" si veda, ad esempio, anche Cortelazzo 2017 che parla di "iperrispecchiamento"). Certamente, in questa testualità troviamo parole che giungono del tutto nuove ad un linguaggio della politica: con Colletti (2013), si può affermare che la lingua della politica "rumoreggia volgare e schietta, diretta e approssimativa".

Leggermente variati o, in alcuni casi, con l'emergenza di ulteriori contenuti, questi manifesti compaiono anche sulla pagina Facebook di Salvini, accompagnati da un conto alla rovescia rispetto all'"Election Day" che rende il messaggio fortemente agganciato alla situazione comunicativa. A partire dal 27

luglio si apre il conteggio con un -60 e, da allora (nell'arco cronologico 27 luglio - 26 agosto ne sono stati contati 23), si susseguono svariati manifesti elettorali in cui è proprio questa indicazione temporale a conquistare il ruolo di *focus* del messaggio: la dimensione dei caratteri numerici, nettamente maggiori rispetto al resto del testo, ne è un chiaro indice. Nel confronto tra i due manifesti (Figg. 9 e 10) si può notare come il secondo sia aderente alla cornice della campagna "Credo" e come, nel tempo, il momento del voto abbia conquistato la scena, realizzando appieno il meccanismo di *embrayage*, presente invece in quello cronologicamente precedente attraverso la deissi personale e temporale, *ora tocca a te!*. Inoltre, si può notare come la testualità del manifesto in Fig. 9, *Ridateci Salvini a difendere i confini*, attraverso lo sfruttamento della rima, rappresenti uno dei pochissimi esempi di slogan in senso stretto e come, per quanto attiene alle dinamiche di enunciazione, riesca a realizzare l'avvicinamento tra mittente e destinatario nel suo configurarsi come un enunciato prodotto dagli elettori.

Fig. 9¹²Fig. 10¹³

La preferenza per un tipo di comunicazione che, di norma, veicola un ordine perentorio in un registro colloquiale, come quello esibito in Fig. 7, accomuna la Lega ad altre componenti del centro-destra, ed era già emersa in

¹² Profilo Facebook di Salvini, 27.07.22.

¹³ Profilo Facebook di Salvini, 17.08.22.

diversi manifesti della Lega stessa, in occasione delle politiche del marzo 2018, come delle europee del maggio 2019, con ricorrente impiego dell'espressione *stop*:

STOP INVASIONE - SALVINI PREMIER - 4 MARZO VOTA - [simbolo Lega/SALVINI-PREMIER]

STOP FORNERO - FATTO! - 26 MAGGIO VOTA - [simbolo Lega/SALVINI-PREMIER]-

STOP INVASIONE - FATTO! - 26 MAGGIO VOTA - [simbolo Lega/SALVINI-PREMIER]-

STOP! - Burocrati - Banchieri - Buonisti - Barconi - EUROPEE · DOMENICA - 26 MAGGIO - VOTA - [simbolo Lega/SALVINI-PREMIER] (Fig.11).

Nel manifesto in Fig. 11 si può notare, tra l'altro, la leggerezza e approssimazione nell'elencare, questa volta, non i prodotti della spesa, ma i mali che a suo modo di vedere affliggono l'Italia, mettendo insieme esseri umani e cose, accomunati dall'averne nel significante l'iniziale /b/ (*barconi* è metonimia reificante e depersonalizzante per 'immigrati clandestini'). Analogo all'impiego di *Stop* è quello dell'espressione *Basta!* riscontrata in un manifesto di questa recente propaganda elettorale, ma antecedente all'avvio della campagna "Credo" (Fig. 12). Da sottolineare come l'uso ricorrente del punto esclamativo negli slogan di Salvini possa essere visto come un'ulteriore traccia dell'iper-presenza del soggetto dell'enunciazione. Come è stato sottolineato (Lala 2018 e 2019), esso svolge la sua funzione non in riferimento a parametri di tipo prosodico o sintattico, come tradizionalmente è stato sostenuto; piuttosto, esso manifesta una "semantica comunicativo-testuale incentrata su valori di tipo interattivo" (Lala 2019: 392) e il suo uso segnala varie forme di coinvolgimento, riconducibili, in termini pragmlinguistici a tre tipi di atti: esclamativi, ottativi e iussivi (Lala 2018: 201), etichettabili anche, rispettivamente, come espressivi, desiderativi, direttivi (Lala 2019: 392). L'atto esclamativo/espressivo - quello realizzato nei testi elettorali di Salvini - implica l'espressione di una valutazione, positiva o negativa, da parte del locutore. Inoltre, in una prospettiva strettamente sociolinguistica, è stato rilevato come il punto esclamativo, nell'esprimere una funzione emotiva, sia tipico di varietà informali dell'italiano; si ritrova, ad esempio, nelle forme di scrittura digitale, ma anche nella scrittura giornalistica e nel linguaggio pubblicitario

(Aprile 2021: 62). Questa caratterizzazione consente di cogliere un ulteriore tratto in comune tra slogan politici e pubblicitari.



Fig. 11¹⁴



Fig. 12¹⁵

Ancora, per quanto concerne l'impiego dei riferimenti alle categorie di mittente e destinatario, si vuole far notare come, sebbene in minor evidenza, collocato proprio in fondo ai manifesti e con caratteri molto piccoli, emerge un ulteriore "io", sia nella formulazione linguistica dell'hashtag #25settebrevotolega (presente nelle due fasi della campagna elettorale), sia nel richiamo al sito [legaonline.it/iocicredo](http://www.legaonline.it/iocicredo) (presente solo a partire dalla campagna "Credo"): si tratta di un "io" per il quale si può ipotizzare l'identificazione di due ruoli: l'io che enuncia, quello di Salvini, che afferma di votare Lega, si fonde con l'io a cui è rivolto il messaggio, l'elettore. Questa possibile interpretazione dell'"io", sulla quale si tornerà più avanti¹⁶, indicherebbe una elaborazione linguistica più sofisticata, dal momento che, di norma, nei manifesti elettorali la formula è *vota X* o *vota per X*. Non a caso, questo tipo di allocuzione si riscontra per lo più in una prima fase della propaganda elettorale,

¹⁴ <http://www.archivisopotpolitici.it/>.

¹⁵ Profilo Facebook di Salvini, 29.07.22.

¹⁶ Par. 3.3.

precedente all'avvio della campagna "Credo", in cui è possibile riconoscere un peso maggiore assegnato al "tu", all'elettore, al quale ci si appella (*Vota Lega!*), adottando nel contempo la tecnica consolidata dello screditamento dell'avversario (*Mandalo a casa*). Il passaggio, dunque, dalla formula *Vota legal!* a quella successiva *VotoLega* (presente nell'hashtag), pur tenendo conto della possibile interpretazione inclusiva ipotizzata sopra, è ulteriore indice di una focalizzazione sul soggetto dell'enunciazione.



**MANDALO A CASA:
IL 25 SETTEMBRE
VOTA LEGA!**



**MAI PATRIMONIALI!
IL 25 SETTEMBRE
VOTA LEGA!**

Fig. 13¹⁷Fig. 14¹⁸

La campagna elettorale di Salvini richiama dunque tutti i tratti tipici del discorso politico dei leader della Lega e di Forza Italia: il leaderismo, la tendenza all'autocelebrazione, l'uso di un linguaggio poco elaborato e facilmente comprensibile (tra gli altri, si veda Ondelli 2017 proprio sul lessico usato da Salvini) e, infine, la spettacolarizzazione. Sin dalle ultime settimane di luglio e nelle prime di agosto, fino alla conseguente esplicitazione del messaggio promozionale, avvenuta il 12 agosto, Salvini ha avviato una campagna pubblicitaria, non ancora politica, in quanto priva di riferimenti a sé o al suo partito, diffondendo in diverse città italiane manifesti con la sola scritta *Credo*, anche attraverso l'uso di cosiddetti camion-vela (Figg. 15-16), alimentando

¹⁷ Profilo Facebook di Salvini, 31.07.22.

¹⁸ Profilo Facebook di Salvini, 14.08.22.

una curiosità crescente sull'identità e il significato nascosti dietro quella singola parola. L'operazione, che può definirsi di vero e proprio marketing, ha raggiunto l'apice della spettacolarizzazione quando il 12 agosto Salvini ha deciso di manifestare la paternità dello slogan, facendo proiettare la parola su edifici e luoghi simbolici di alcune città italiane: il porto di Lampedusa, l'Agenzia delle Entrate di Roma, la stazione centrale di Milano, la sede dell'Inps di Roma, con l'obiettivo di focalizzare l'attenzione sui temi cari alla sua politica: lo stop all'immigrazione, la riforma delle pensioni e del sistema fiscale, la sicurezza nelle città.



Fig. 15



Fig. 16

Contemporaneamente, Salvini ha diffuso lo slogan sui propri canali “social”, legandolo definitivamente al suo nome, e ha pubblicato sul suo canale YouTube un video che riprende le proiezioni, accompagnato da un audio in cui si susseguono, prima la voce dell'attore Pino Insegno e poi quella dello stesso Salvini, che, con toni sempre fortemente auto-referenziali, sostiene di aver cominciato a fare politica da giovanissimo, “perché – come tanti – credevo e ancor più oggi *credo* in valori e ideali, nella lealtà e nell'onestà, in un futuro migliore”. Quest'operazione di marketing riecheggia, sebbene con una piccola variazione diamesica, l'analoga campagna inaugurata da Berlusconi nel 2001, quando, sempre con grande anticipo rispetto all'inizio della vera e propria campagna elettorale, il leader di Forza Italia aveva fatto ricoprire intere facciate di palazzi e di impalcature, originariamente destinate ad accogliere cartelloni pubblicitari di tipo commerciale, con maxi-manifesti del proprio partito, suscitando peraltro anche reazioni/azioni legali da parte dei condomini i cui edifici erano stati interessati dall'affissione (Ottaviano 2004: 287). In un'epoca in cui i *media* e in particolare la televisione sembravano

aver monopolizzato i luoghi dello scontro elettorale, il manifesto elettorale riaffermava la sua forza, con l'amplificazione delle dimensioni canoniche per questo tipo di manifesto, determinando "un'attenzione motivata dal gigantismo e dall'onnipresenza dei manifesti" (Sergio 2008: 10). Il gigantismo realizzato e perseguito da Berlusconi continua, con le variazioni segnalate, nell'operazione pubblicitaria di Salvini, che pare assumere tratti tipici di un atteggiamento megalomane.

Ma l'operazione non termina qui: il giorno successivo all'avvio spettacolarizzato della campagna "Credo" parte su Twitter, in diversi account di supporto a Salvini¹⁹, un'ulteriore iniziativa di marketing pubblicitario che conferma le osservazioni fatte fin qui. La formula *Credo* si rafforza in *Io ci #Credo*, attraverso una serie di manifesti in formato webcard in cui il "Credo" di Salvini contrasta con quello di altri esponenti politici, in un quadro che mostra chiara la dinamica di una deissi pronominale che vede l'"io" contrapposto al "lui/lei". Le testualità sono molteplici (a seguire sono riprodotti solo un paio di manifesti, Figg. 17-18) e vi è *variatio* anche nell'associazione di un medesimo slogan a volti di politici diversi:

LUI NON - CREDE NELLA - LIBERTÀ - [simbolo Lega/SALVINI-PREMIER] - *Io ci #Credo*

LUI NON CREDE - NELLA DIFESA - DEI CONFINI - [simbolo Lega/SALVINI-PREMIER] - *Io ci #Credo*²⁰

LUI NON - CREDE NELLA - PACE FISCALE- [simbolo Lega/SALVINI-PREMIER] - *Io ci #Credo*

LEI NON - CREDE NELLA - SICUREZZA - [simbolo Lega/SALVINI-PREMIER] - *Io ci #Credo*

LEI NON - CREDE NELLA - FAMIGLIA- [simbolo Lega/SALVINI-PREMIER] - *Io ci #Credo*

LEI NON - CREDE NELLA - PACE FISCALE - [simbolo Lega/SALVINI-PREMIER] - *Io ci #Credo*.

¹⁹ Tra cui: Noi con Salvini, @Noiconsalvini, e Lega - Salvini Premier, @LegaSalvini.

²⁰ Ad esempio, questo slogan è associato ai volti di Beppe Sala, Matteo Orfini, Roberto Saviano.



Fig. 17

Fig. 18²¹

3.2. Le scelte del PD: dalla terza alla seconda persona

Rispetto alla posizione appena analizzata, in cui emerge l'insistenza sul soggetto dell'enunciazione, con una forte affermazione dell'"io", si può collocare come antitetica la scelta del Partito Democratico, almeno nella prima fase della campagna elettorale. Non è un caso che siamo a parlare di un partito e non di un leader: la posizione del PD si colloca agli antipodi rispetto alle posizioni tipiche di buona parte della politica italiana dalla cosiddetta Seconda Repubblica in poi, richiamandosi invece alla tradizione classica della sinistra, per quanto il PD se ne possa essere allontanato per le scelte politiche. Facendo riferimento a diversi esponenti della sinistra, da quella più radicale e classica a quella via via sempre più moderata, si può riconoscere uno stile non dialogico e orientato invece alla realtà di cui si parla: un linguaggio "dominato dall'uso della forma impersonale, con marcata assenza di toni interattivi ed emozionali verso l'uditorio" (Desideri 2011, in relazione allo stile di Berlinguer), caratterizzato da uno stile didattico, nel quale domina la III persona, e che "mira a realizzare un processo di identificazione del ricevente con i contenuti del messaggio" (Santulli 2005: 119, in riferimento allo stile di Prodi, contrapposto a quello di Berlusconi).

²¹ Entrambi i manifesti sono apparsi sul profilo Twitter della Lega (Lega - Salvini Premier, @LegaSalvini) il 13.8.22.

Queste osservazioni sono perfettamente estendibili all'analisi dello stile adottato dal PD nella prima versione della campagna elettorale (a titolo esemplificativo si riportano i due manifesti in Figg. 19-20), quella il cui slogan è *Vincono le idee*, anche in questo caso via via arricchito di contenuti, con esplicitazione delle idee in materia di giustizia sociale, diritti civili, economia, ambiente, ecc.:



Fig. 19



Fig. 20

UN MESE - DI STIPENDIO IN PIÙ - Giù le tasse sul lavoro, su i salari. - [simbolo PD] - Vincono le idee. PARTECIPA - E VINCIAMO - INSIEME. - Scopri come su - partitodemocratico.it

AVANTI - SUI DIRITTI CIVILI - Senza paura, per una società di persone libere. - [simbolo PD] Vincono le idee. PARTECIPA - E VINCIAMO - INSIEME. - Scopri come su - partitodemocratico.it

PARITÀ SALARIALE - TRA DONNE E UOMINI - Stesso lavoro, stesse retribuzioni. - [simbolo PD] - Vincono le idee. PARTECIPA - E VINCIAMO - INSIEME. - Scopri come su - partitodemocratico.it

AFFITTI PIÙ BASSI PER I GIOVANI - Contributo di 2000€ l'anno per studenti e lavoratori under 35. - [simbolo PD] - Vincono le idee. PARTECIPA - E VINCIAMO - INSIEME. - Scopri come su - partitodemocratico.it

PIÙ MEDICI DI FAMIGLIA - Piano straordinario per stare vicino a chi ha bisogno di cura. - [simbolo PD] - Vincono le idee. PARTECIPA - E VINCIAMO - INSIEME. - Scopri come su - partitodemocratico.it

CHI STUDIA - IN ITALIA È ITALIANO - Ius scholae per i bambini che vanno a scuola con i nostri figli. [simbolo PD] - Vincono le idee. PARTECIPA - E VINCIAMO - INSIEME. - Scopri come su - partitodemocratico.it

PRIMA - L'AMBIENTE - Per la salute, per l'economia, per i nostri figli. - [simbolo PD] - Vincono le idee. PARTECIPA - E VINCIAMO - INSIEME. - Scopri come su - partitodemocratico.it

I BAMBINI - SONO TUTTI UGUALI - Scuola dell'infanzia obbligatoria e gratuita. - [simbolo PD] - Vincono le idee. PARTECIPA - E VINCIAMO INSIEME. - Scopri come - su partitodemocratico.it

LA CASA - È UN DIRITTO - 500.00 nuovi alloggi a canone concordato in 10 anni. - [simbolo PD] - Vincono le idee. PARTECIPA - E VINCIAMO INSIEME. - Scopri come - su partitodemocratico.it

IL LAVORO - È DIGNITÀ - Salario minimo, - lotta alla precarietà - e al lavoro nero. - [simbolo PD] - Vincono le idee. PARTECIPA - E VINCIAMO INSIEME. - Scopri come - su partitodemocratico.it

ITALIA - RINNOVABILE. - Zero burocrazia per aziende green, più lavoro, bollette più basse. - [simbolo PD] - Vincono le idee. PARTECIPA - E VINCIAMO INSIEME. - Scopri come - su partitodemocratico.it

MAI PIÙ - FINTI STAGE. - Nel mercato del lavoro solo apprendistato retribuito. - [simbolo PD] - Vincono le idee. PARTECIPA - E VINCIAMO INSIEME. - Scopri come - su partitodemocratico.it

In questi manifesti, che hanno avuto una discreta circolazione, si possono notare tutti i caratteri già rilevati nei discorsi dei politici di sinistra, confermando il fatto che gli slogan rappresentino una sorta di condensato del discorso vero e proprio. Le strutturazioni linguistiche sono deitticamente sganciate dal mittente dell'enunciazione, né si appellano a un destinatario, l'elettore; sono cioè svincolate dalla dimensione dell'interazione comunicativa, presentando un dominio della III persona, cioè dei fatti, proposti come una realtà oggettivizzata. In termini pragmatici, c'è una dominanza della funzione referenziale, mentre sono ridotte tanto la funzione espressiva quanto quella conativa. Unico tratto che manifesta una deissi di persona, per altro in una modalità inclusiva, è l'impiego dell'aggettivo possessivo *nostri* nel manifesto che recita *Chi studia in Italia è italiano. Ius scholae per i bambini che vanno a scuola con i nostri figli* e in quello di vocazione ambientalista, *Prima l'ambiente. Per la salute, per l'economia, per i nostri figli*. Peraltro, si noti che nel primo testo, a dispetto dell'obiettivo inclusivo della proposta politica, viene di fatto a crearsi una contrapposizione tra una sfera del "noi" (*i nostri figli*) e dell'"altro", la III persona (*i bambini che vanno a scuola*). Si vuole far notare a questo proposito come, sebbene questo lavoro si concentri esclusivamente su tre forze politiche, il ricorso all'uso dell'aggettivo possessivo di I plurale abbia rappresen-

tato per altri schieramenti politici la scelta per veicolare inclusività, come, ad esempio, nei messaggi che seguono, nei quali *il nostro presidente e il nostro Paese* alludono ad una dimensione pienamente collettiva:

GIUSEPPE - CONTE - *Il nostro Presidente* - DOMENICA 25 SETTEMBRE - SCEGLI IL - MOVIMENTO 5 STELLE - [Simbolo Movimento 5 Stelle]

Dobbiamo totalmente - cambiare le politiche - energetiche, ambientali - e industriali del nostro Paese - LUIGI DE MAGISTRIS - [Simbolo Unione Popolare con de Magistris].

Tornando ai manifesti della campagna “Vincono le idee”, si può notare come il ridotto utilizzo dei meccanismi di riferimento personale apra la strada, come già detto, ad un maggior peso assegnato alla funzione referenziale, e si accompagni a qualche elemento di sfruttamento della funzione poetica: strutture di antitesi con *giù* e *su* (*giù le tasse, su i salari*), strutture di ripetizione con *stesso/stesse* (*Stesso lavoro, stesse retribuzioni*) e *per* (*per la salute, per l'economia, per i nostri figli*) che, non a caso, avvicinano queste messaggi agli slogan in senso stretto.

L'assenza di riferimenti ai ruoli dell'enunciazione in questa prima versione della campagna è inoltre ribadita dall'assenza dell'immagine del leader del PD: la vittoria alle elezioni è del tutto priva di riferimenti alle persone; non è il partito, tanto meno il suo leader, che vincerebbero le elezioni, ma le idee, con la loro forza. Si noti il forte contrasto con l'immagine di Salvini con il pugno chiuso o di Giorgia Meloni con le dita disposte a V, in segno di vittoria. L'accusa di “autoreferenzialità”, spesso mossa al PD, probabilmente va intesa anche nel senso di una mancanza di comunicatività, cioè di capacità di parlare agli elettori, di proporsi ad essi come persone e di farli sentire coinvolti.

Gli unici elementi di dialogicità, in posizione marginale, sono realizzati nel riquadro in giallo, in cui si legge finalmente un appello all'elettore: *Partecipa e vinciamo insieme. Scopri come su **partitodemocratico.it***; emergono sia la I persona che una I persona plurale, chiaramente inclusiva (*vinciamo insieme*). A ben vedere, però, non si tratta di un appello al voto, previsto e, diremmo, anche legittimo in un manifesto elettorale, ma di un appello ad una generica partecipazione, quasi che l'operazione di voto, da altre parti politiche invocata e ricordata proprio attraverso l'immagine dell'apposizione della X sul simbolo del partito o sul nome del premier, fosse stata tabuizzata. Peraltro, non

si può evitare di notare come tutta la formulazione linguistica evochi una dimensione altra rispetto al voto e vicina a quella della partecipazione a un concorso a premi, che condivide con la prima la modalità (la partecipazione) e l'obiettivo (la vincita). Anche l'invito a scoprire come vincere e l'immagine di un QRcode evocano tutti i connotati della partecipazione a un concorso a premi (nel manifesto in Fig. 21 è più in evidenza il contenuto del riquadro in giallo, anche se manca il QrCode, presente negli altri manifesti):



Fig. 21



Fig. 22

Sta di fatto che, come noto, questa strategia oggettivizzante non si è rivelata vincente, tanto che, a partire dai primi di settembre, il PD, allarmato dai sondaggi che davano FdI come partito nettamente vincente, ha deciso di cambiare politica comunicativa, passando alla seconda fase della campagna elettorale, contraddistinta dallo slogan *Scegli*. In questa nuova fase, emerge quindi in primo piano il destinatario che è chiamato ad agire, operando una scelta, che viene sempre presentata come alternativa ad un'altra. In questa strategia comunicativa, quindi, anche senza nominare direttamente l'avversario politico, si riconosce nettamente la scelta di presentare argomentazioni "contro", sfruttando la tecnica dell'implicito (su questo si vedano, in particolare, i lavori di Lombardi Vallauri del 2015 e 2019, e quello con Masia del 2016). In questo caso si può sottolineare come lo slogan politico operi secondo gli stessi meccanismi del messaggio pubblicitario, mettendo in atto in particolare la strategia della pubblicità comparativa (si veda al riguardo De Santis 2019: 365). Rispetto alla prima fase della campagna elettorale, il linguaggio è ancora più asciutto, ridotto ai soli nuclei comunicativi. Compare la foto del leader

(ma si noti che l'immagine è volutamente tagliata, quasi a volerne ridimensionare il peso nell'economia generale del manifesto); il leader non enuncia nulla di sé, ma si rivolge al destinatario in modo secco, con una formula che ancora una volta mostra una tabuizzazione del voto. Anche in questo caso, sulla base di una stessa struttura (esempi in Figg. 23-24), si sono sviluppate molteplici variazioni testuali che vanno dalla contrapposizione tra due sole parole a quella tra due sintagmi nominali o preposizionali, con un isolato caso in cui si sviluppa una predicazione più articolata (*per chi lavora*):

DISCRIMINAZIONI / DIRITTI - [simbolo PD] Scegli.

PIÙ - CONDONI - PER GLI EVASORI/ MENO - TASSE - PER CHI LAVORA - [simbolo PD] Scegli.

CON - PUTIN / CON - L'EUROPA - [simbolo PD] Scegli.

NO VAX / SCIENZA - E VACCINI - [simbolo PD] Scegli.

COMBUSTIBILI - FOSSILI / ENERGIE - RINNOVABILI - [simbolo PD] Scegli.

LAVORO - SOTTO - PAGATO / SALARIO - MINIMO - [simbolo PD] Scegli.

SANITÀ A PAGAMENTO - PER POCHI / SANITÀ GRATUITA - PER TUTTI - [simbolo PD] Scegli.



Fig. 23



Fig. 24

Rispetto alla prima fase, c'è il proposito esplicito di Letta di coinvolgere l'elettore, chiamandolo a prendere posizione (e si noti che il punto fermo amplifica l'incisività dell'invito), presentando l'alternativa con contrapposizione di opzioni politiche collocate in zone colorate di nero (l'altro) o rosso (il

sé). La tecnica della classificazione e comunicazione dei contenuti in termini “binari” (Cosenza 2018: 5) in questa versione della campagna appare esasperata, proprio in virtù del fatto che i messaggi sono essenziali, presentando solo temi, senza predicazione e senza sintassi alcuna; come lo stesso Enrico Letta afferma “è il momento della polarizzazione, delle parole nette e di una comunicazione anche brutale della posta in gioco. Bisogna dare la sveglia agli italiani”²². Si noti la differenza rispetto alla strategia comunicativa di Salvini che ha lo stesso intento di giocare sull’*argumentum ad personam* (Figg. 17-18), ma con manifestazione esplicita degli attori della comunicazione.

Questa nuova campagna non è particolarmente originale né nella strategia comunicativa di appellarsi direttamente all’elettore, con un canonico imperativo di II persona (*scegli, decidi, vota, scrivi*), né nella selezione del verbo *scegliere* in particolare, usato già in passato tanto dalla sinistra del PD, quanto dalla destra²³. Può essere utile ricordare i manifesti delle elezioni politiche del 4 marzo 2018, in cui, peraltro, i contenuti veicolati appaiono sostanzialmente gli stessi del 25 settembre 2022, che consentono di fare un confronto con la nuova versione della campagna “Scegli”: in quei manifesti emergeva un maggiore aggancio alla situazione comunicativa, con l’indicazione tanto della data quanto della giustapposizione della X sul simbolo del partito; eppure, anche allora, per certi aspetti, l’indicazione di voto era spostata dal partito alla tematica di cui il partito si faceva portavoce: *Vota il lavoro e non vota il PD*:

4 marzo - VOTA IL LAVORO - SCEGLI IL PD - [simbolo PD barrato] - *partitodemocratico.it*

4 marzo - VOTA L'EUROPA - SCEGLI IL PD - [simbolo PD barrato] - *partitodemocratico.it*

4 marzo - VOTA LA SCIENZA - SCEGLI IL PD - [simbolo PD barrato] - *partitodemocratico.it*

4 marzo - VOTA L'AMBIENTE - SCEGLI IL PD - [simbolo PD barrato] - *partitodemocratico.it*

4 marzo - VOTA LA CULTURA - SCEGLI IL PD - [simbolo PD barrato] - *partitodemocratico.it*

4 marzo - VOTA I DIRITTI - SCEGLI IL PD - [simbolo PD barrato] - *partitodemocratico.it*.

²² Parole tratte da un’intervista per Repubblica, https://www.repubblica.it/politica/2022/08/25/news/letta_pd_elezioni-362824944/.

²³ L’imperativo *scegli* è presente ad esempio in un manifesto di Alleanza Nazionale per le politiche del 2006: SCEGLI - [simbolo AN barrato] - *L’Italia. Una passione in comune*.



Fig. 25

Fig. 26²⁴

Il confronto con la campagna del 2018 offre nuove argomentazioni per sostenere che se, nelle intenzioni del leader del PD, questa nuova versione della campagna si prefiggeva l'obiettivo di fare maggiormente appello agli elettori, di fatto, la sua forza pragmatica continua ad essere molto debole. Si noti che, accanto all'emergere dell'immagine del leader, l'unica ulteriore differenza con la prima fase della campagna è rappresentata dall'appello al destinatario, chiamato in causa con un imperativo, mentre l'evento del voto continua ad essere oscurato, rendendo di fatto il messaggio svincolato dalla situazione comunicativa, secondo una tecnica di *débrayage*: come nei testi della prima fase manca alcun tipo di riferimento temporale al giorno delle elezioni (ancora una volta in netto contrasto con il conto alla rovescia rispetto al giorno delle votazioni osservato nei manifesti di Salvini). Di conseguenza, il coinvolgimento dell'elettore non appare realmente accresciuto, ma, se possibile, perfino ridotto dal venir meno di parole che, sebbene con l'interferenza semantica evidenziata sopra, evocano la dimensione del voto: *vincono, partecipa* (contro l'astensionismo), *vinciamo*.

Inoltre, si accentua la separazione tra il leader politico (che si manifesta solo con l'immagine) e l'elettore, che è chiamato, quasi da solo, ad assumersi la responsabilità

²⁴ Manifesti tratti dall'archivio <http://www.archivisopotpolitici.it/archivio>.

delle conseguenze delle elezioni, facendo la “scelta” giusta. In definitiva, ciò che manca è la dimensione del “noi” che accomuna partito o leader e elettore.

Sebbene il leader del PD sfrutti in misura limitata i canali “social” Facebook e Twitter a fini elettorali e propagandistici (pubblicando nel periodo tra il 21 luglio e il 25 settembre un numero di post al giorno decisamente inferiore a quelli pubblicati da Salvini e Meloni), un esame di questi canali conferma alcune delle osservazioni sopra proposte. I contenuti pubblicati raramente fanno esplicito riferimento all’appuntamento elettorale o formulano un diretto invito al voto. Infatti, diversamente da quanto osservato nel caso di Salvini e di quanto si evidenzierà per Meloni, pochissimi sono i manifesti elettorali pubblicati da Letta su questi *media* e si tratta quasi esclusivamente di quelli già presenti sul sito ufficiale del partito. Come aspetto pertinente rispetto alla manifestazione della deissi di persona, si può rilevare come nel commento in un post del 18 settembre sull’inondazione verificatasi nelle Marche - che pur non configurandosi come manifesto elettorale è pubblicato in piena propaganda elettorale (Fig. 27) - emergano due diverse accezioni di “noi” come “persona politica”: la prima occorrenza della I forma plurale (*siamo*) fa riferimento ad un “noi” collettivo che designa tutti i politici, mentre la seconda occorrenza (*facciamo*) si riferisce ad un “noi” esclusivo della destra: *Non è vero che siamo tutti uguali. La destra nega i cambiamenti climatici, noi facciamo proposte per risolvere i problemi.*



Fig. 27

Fig. 28²⁵

²⁵ Facebook Letta, 22.09.22.

Analizzando nel dettaglio il profilo Facebook, si osserva che il richiamo esplicito al voto è presente esclusivamente in due hashtag, #VotoPD e #25settembrevotoPD, che ricorrono a chiusura di diversi dei testi introduttivi di accompagnamento ai contenuti pubblicati da Enrico Letta. E nel manifesto in Fig. 28 si può cogliere nella forma verbale *voto* la presenza di un “io” ambivalente che, come già notato per Salvini, rappresenta tanto l’estensore del messaggio, in questo caso Letta, quanto l’elettore, mentre la particella pronominale *ci* (*ci ha dato una spinta enorme in questa campagna*) sembra far riferimento ad un “noi” esclusivo (‘noi del PD’) se non ad un “io maiestatico” riferito al leader del PD.

Rispetto all’opzione così marcata del PD di evitare meccanismi di inclusione linguistica, si può osservare come l’altra forza di sinistra, più radicale, rappresentata dall’Alleanza Verdi e Sinistra, nella sua scelta per questa campagna elettorale, pur ricalcando caratteristiche che, come abbiamo notato, sono state e continuano ad essere tipiche della comunicazione politica di sinistra - mancanza dell’immagine del leader, cancellazione dell’“io” e del “tu”, con conseguente effetto di rendere oggettivati i contenuti del messaggio, omissione dell’indicazione del voto e della data delle elezioni -, realizza un discreto tasso di inclusività attraverso l’impiego di un esortativo di I plurale, *facciamolo*. Da notare che la forma verbale - come già fatto osservare a proposito dei manifesti del PD - non è seguita da un punto esclamativo (in grande contrasto con le scelte di Salvini) ma da un punto fermo, con l’obiettivo di dare maggiore incisività all’invito, che sembra configurarsi come qualcosa di realizzabile. Proprio questa forma verbale, che evoca il “noi”, rappresenta lo slogan della formazione politica, che la propone in molteplici manifesti:



Fig. 29



Fig. 30

Sulla base di questa struttura, le testualità, come in altri casi, sono molteplici:

AGENDA VERDE, - L'UNICA CHE - VOGLIAMO - Legge per il clima e - Transizione ecologica - Facciamolo. [simbolo Alleanza Verdi Sinistra] - www.verdisinistra.it

ISTRUZIONE - COMPLETAMENTE - GRATUITA - dalla culla all'università - Facciamolo. [simbolo Alleanza Verdi e Sinistra] - www.verdisinistra.it

SALARIO MINIMO - di almeno 10€ l'ora - Facciamolo. [simbolo Alleanza Verdi e Sinistra] - www.verdisinistra.it

TRASPORTO - PUBBLICO - GRATUITO - per almeno 10 mesi all'anno - Facciamolo. [simbolo Alleanza Verdi e Sinistra] - www.verdisinistra.it

LEGGE - SUL CLIMA - IN 100 GIORNI - Facciamolo. - [simbolo Alleanza Verdi e Sinistra] - www.verdisinistra.it

STOP CARO - BOLLETTE - Extraprofitti agli italiani - Facciamolo. [simbolo Alleanza Verdi e Sinistra] - www.verdisinistra.it.

Si noti come in alcuni di questi manifesti ci sia la ricerca di strutturazioni linguistiche che sfruttano la funzione poetica, tendenzialmente assente negli altri manifesti analizzati: nel testo *L'Italia è donna* (Fig. 30), volendo veicolare contenuti politici che vanno nella direzione di una attenzione al mondo femminile, si gioca sul fatto che il nome *Italia* sia femminile e ancora (Fig. 29), ma molto meno ricercata, e simile ai meccanismi degli slogan pubblicitari, emerge attraverso il sintagma *agenda verde* la tecnica dell'impiego di una stessa parola, *agenda*, in senso letterale e figurato, vale a dire come programma di attività e impegni in materia ecologica e come taccuino di colore verde rappresentato nell'immagine. Peraltro, in questa stessa testualità il valore di inclusività è ulteriormente rafforzato da un'altra forma verbale di prima plurale: *vogliamo*.

3.3. Giorgia Meloni e la molteplicità delle persone

A differenza delle posizioni della Lega e del PD, polarizzate tra una scelta dominante in favore dell'"io" o del "tu", i messaggi elettorali di Giorgia Meloni mostrano una pluralità di opzioni, di volta in volta incentrate sui diversi attori dell'enunciazione, ma con una forte propensione a rappresentare un senso di unità tra sé (o il proprio partito) e l'elettorato; si fa notare sin da adesso che, rispetto alle strategie di propaganda elettorale delle altre forze politiche fin qui analizzate, la campagna di Giorgia Meloni si sviluppa maggiormente sui canali social.

La formula *Pronti*, come si mostrerà, rappresenta la parola chiave intorno alla quale si declinano diversi messaggi elettorali della leader di Fdi: può presentarsi estremamente sintetica (in alcuni casi anche da sola: *Pronti* (Fig. 32), o arricchita nello slogan che è circolato maggiormente, sotto forma di cartellone, nella campagna elettorale: *Pronti a risollevere l'Italia* (Fig. 31). In entrambi i casi, lo slogan risulta volutamente ambiguo, nella misura in cui può trasmettere diversi sensi: può essere letto come un'affermazione dell'essere pronti da parte dei politici di Fratelli d'Italia, o può essere interpretato come una sorta di invito, una richiesta rivolta agli elettori ad essere pronti, a contribuire, attraverso il loro voto, alla realizzazione dell'obiettivo (*risollevere l'Italia*), o, ancora, può riferirsi congiuntamente a parte politica ed elettori rappresentati come un tutt'uno. Infatti, la forma plurale ha insita una indeterminatazza della referenza e permette, quindi, un riferimento inclusivo/esclusivo di mittente e ricevente. Il riferimento personale in questo caso non passa attraverso l'indicazione del pronome, né attraverso la marca di persona nel verbo, che è assente (esplicitato invece, come si analizzerà, in *Siamo pronti*, accanto a *Sono pronta* e *siete pronti*), ma semplicemente attraverso il plurale dell'aggettivo predicativo.



Fig. 31

Fig. 32²⁶

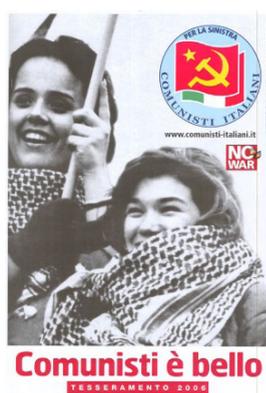
Questo modo di realizzare linguisticamente un senso di unione e condivisione risulta più sottile e meno immediatamente percepibile dell'impiego del "noi" inclusivo, eppure è di grande forza, tanto che è stato utilizzato in passato anche da altre forze politiche. A titolo esemplificativo, si mostrano due

²⁶ Profilo Twitter di Fratelli d'Italia, @Fratelliditalia, 17.08.22.

manifesti che si inscrivono, il primo, in un contesto elettorale (2003) e il secondo, in una campagna di tesseramento al partito (2006). In entrambi casi, il senso di appartenenza è veicolato esclusivamente dal plurale dei sostantivi/aggettivi, tanto più che la sintassi dei due slogan è molto diversa:



Fig. 33

Fig. 34²⁷

In realtà lo slogan utilizzato da Meloni, pur con una struttura predicativa più piena, *Siamo pronti*, era comparso già il 21 luglio (su Facebook e su Twitter, Fig. 35), subito prima dell'avvio vero e proprio della campagna elettorale (le elezioni sono state indette con DPR del 21/7/22), proponendo un “noi” decisamente inclusivo. Questo manifesto appare connotato da caratteri propri della destra storica, facendo appello ai patrioti, con una formula fortemente conativa, *Avanti, Patrioti*, e richiamando a livello di immagine i colori della bandiera nazionale (in analogia con la scelta di Salvini) e il simbolo della vittoria nel gesto della Meloni. L'esortazione *Avanti*, con la componente semantica di deissi spaziale, contribuisce ad ancorare il discorso all'*hic et nunc* dell'enunciazione, insieme alla deissi personale nel verbo, *Siamo*, il quale, inoltre, implica anche un aggancio temporale, '(adesso) siamo pronti': *essere pronti* è risultativo, designa il punto di arrivo di un processo, è momentaneo, non durativo, e rappresenta anche una proiezione temporale, coincidendo

²⁷ Manifesti tratti dalla banca dati manifestopolitici.it.

con il punto di inizio di un'azione potenziale. Questo manifesto, pubblicato precocemente, consente di cogliere un aspetto dell'evoluzione del simbolo del partito: si può infatti notare come il logo di Fratelli d'Italia non si presenti ancora nella veste che sarà definitivamente messa a punto per le elezioni (Fig. 36), nella quale si deciderà di porre ad esponente il nome completo di Giorgia Meloni, senza ulteriori qualificazioni (rispetto a quelli di Salvini e Berlusconi, etichettati, rispettivamente, come *premier* e *Presidente*), ma con caratteri di dimensioni marcatamente maggiori rispetto a quelli utilizzati per il nome del partito, rendendo di fatto quest'ultimo inglobato in quello della leader. Questa osservazione ci porta a rilevare un tratto di personalizzazione anche nella comunicazione elettorale di Fratelli d'Italia; tuttavia, a questo punto, risulta tanto più importante sottolineare come, nel confronto con il tipo di comunicazione di Salvini, la strategia comunicativa di Giorgia Meloni risulti non così contraddistinta da toni personalistici proprio grazie alla scelta di puntare su scelte linguistiche che si sviluppano intorno al polo della pluralità, inclusiva o esclusiva che sia.

Fig. 35²⁸

Fig. 36

In altri luoghi, infatti, la scelta di Giorgia Meloni di utilizzare il plurale (*Pronti*) risulta funzionale a veicolare un valore di polarizzazione del “noi” esclusivo (‘noi come parte politica’) in contrapposizione dialogica con il “voi” (gli elettori). Nello sviluppo della campagna, lo schema dello slogan va riempien-

²⁸ Profili Twitter e Facebook di Fratelli d'Italia, 21.07.22.

dosi di contenuti più circostanziati che implicano come soggetto dell'agire la parte politica: 'noi di Fratelli d'Italia siamo pronti a governare e a realizzare tutta una serie di propositi', che compaiono via via in diversi manifesti (Fig. 37-38), creati a partire da video di propaganda presenti sui canali social di Giorgia Meloni:

25 settembre 22 - [simbolo FdI] - Pronti - a dare un sostegno - concreto alle famiglie

25 settembre 22 - [simbolo FdI] - Pronti - a tutelare la casa - degli italiani

25 settembre 22 - [simbolo FdI] - Pronti - a ridare forza - alla volontà popolare

25 settembre 22 - [simbolo FdI] - Pronti - a ridare sicurezza - alle nostre città (ma si noti che nostre è inclusivo)

25 settembre 22 - [simbolo FdI] - Pronti - a proteggere i lavoratori autonomi

25 settembre 22 - [simbolo FdI] - Pronti - a difendere chi fa impresa - rispettando le regole.



Fig. 37²⁹



Fig. 38³⁰

La formulazione linguistica piena, *Noi siamo pronti*, con esplicitazione del *noi*, compare in un manifesto (Fig. 39) in cui il pronome si manifesta ancora con la sua valenza di esclusivo, facendo riferimento alla parte politica, e non si esclude anche l'interpretazione di un *noi* "maiestatico", dal momento che, a livello di

²⁹ Profilo Facebook di Meloni, 26.08.22.

³⁰ Profilo Facebook di Meloni, 24.08.22.

immagine compare sempre e solo la leader di Fratelli d'Italia. Tuttavia, analizzando attentamente l'intero testo, *Per l'Italia. Ecco il programma del Centrodestra per il Governo dei Patrioti. Noi siamo pronti*, si coglie un altro punto di ambivalenza: il *Governo dei Patrioti* può essere inteso nel senso di 'dei patrioti che governano' o 'dei patrioti che sono governati' in virtù della indeterminatezza semantica del sintagma preposizionale retto da *di*, che dà origine al genitivo soggettivo o oggettivo.



Fig. 39³¹

A conferma del pieno sfruttamento dell'ambiguità della formula *Pronti*, fulcro della campagna elettorale, si può notare come nel tempo si manifesti anche divaricazione dei riferimenti deittici, con l'effetto di una evidenziazione della struttura dialogica:

Io sono pronta a governare questa Nazione. - Se voi siete pronti - a un Governo libero, - il 25 settembre - mettete una croce - sul simbolo - di Fratelli d'Italia - Giorgia Meloni - [simbolo FdI barrato] (Fig. 40);

Sì, sono pronta, - se voi siete - pronti ad avere - un Governo - di persone libere, - che non si fanno - comprare e che - vogliono restituire - l'orgoglio - a questa Nazione - [simbolo FdI barrato] (Fig. 41).

In queste strutturazioni il 'noi siamo pronti a governare/risolvere l'Italia' si declina come *io sono pronta a governare*, da una parte, e *voi siete pronti a un Governo*, dall'altra: emerge dunque il pronome di I singolare, ma assolutamente privo

³¹ Profilo Twitter di Meloni, 12.08.22.

della forza accentratrice presente nelle testualità di Salvini (o in passato nei messaggi di Berlusconi), sebbene il primo manifesto presenti la firma *Giorgia Meloni*: l'“io” si relaziona sempre al “voi” attraverso il periodo ipotetico che pone la condizione, il fondamento del “patto” tra la Meloni e i suoi elettori³². Nella seconda testualità la forza assertiva è marcata dal Sì iniziale che dà all'enunciato la struttura di una risposta ad una domanda, implicando l'interlocuzione con gli elettori, comunque attivata dall'impiego del pronome *voi*; peraltro, la forza pragmatilinguistica è incrementata dall'uso del dimostrativo (*questa Nazione*). Nel primo testo, è da notare, in contrapposizione con l'atteggiamento emerso nei manifesti del PD, l'invito puntuale all'operazione di voto, rivolta a tutti i potenziali elettori: *mettete una croce sul simbolo di Fratelli d'Italia* (contenuto ribadito nell'immagine).

Fig. 40³³Fig. 41³⁴

Il “noi” esclusivo emerge in altri luoghi, sempre però contestualmente a valori semantici di inclusività: ad esempio, nel testo *La corsa non è finita. Ma gli italiani sono con noi. Avanti così!* (Fig. 42), si può notare come, accanto alla evidente manifestazione di un “noi” esclusivo (*gli italiani sono con noi*), si realizzi un senso di unità con l'elettorato, veicolato dall'implicito che la corsa è stata comune e che è richiesto un ulteriore passo (*avanti così!*), ancora una volta in una cornice

³² Sui profili Facebook e Twitter di Meloni compare il 01.09.22.

³³ Facebook Meloni, 01.09.22.

³⁴ Twitter e Facebook Meloni, 16.09.22.

di incitamento verso un'azione congiunta. Anche nell'altro messaggio (Fig. 43) si può osservare il gioco di contrapposizione tra un "noi" inclusivo (*abbiamo bisogno di un governo di persone*), come se la stessa Giorgia Meloni fosse nella condizione di essere governata, ed un "noi" esclusivo, (*noi non ci facciamo intimorire, ricattare o comprare*), tra l'altro messo in maggiore evidenza attraverso l'impiego di caratteri di corpo maggiore e la scelta del colore giallo.

Fig. 42³⁵Fig. 43³⁶

Ma è sicuramente più frequente l'impiego, da parte della leader di Fratelli d'Italia, di un "noi" inclusivo che mira a consolidare, come già argomentato, un senso di unione e condivisione, come emerge dalle immagini che mostrano Giorgia Meloni gestalticamente come "figura" in primo piano, sullo "sfondo" della folla (Fig.44). Si può notare ancora una volta lo sfruttamento degli operatori deittici in formulazioni in cui questi sono associati a espressioni verbali che si riferiscono ad azioni da realizzare nell'immediato, *ora riempiamo*, rispetto ad azioni già effettuate, *abbiamo riempito*. Nello stesso testo si manifesta anche la funzione poetica nella ripetizione modulata del verbo *riempire*.

³⁵ Twitter e Facebook Meloni, 6.09.22.

³⁶ Twitter e Facebook Meloni, 25.08.22.

Fig. 44³⁷Fig. 45³⁸

Nel manifesto in Fig. 45, l'insistenza sul "noi inclusivo" si riconosce nella forma verbale di I plurale, rafforzata dall'uso dell'avverbio *insieme*. Inoltre, la forma imperativo-esortativa invita ad un'azione immediata e congiunta. L'uso del punto esclamativo, come ricordato sopra³⁹, può assolvere a funzioni comunicativo-testuali diverse, ma sempre pragmaticamente molto forti ed efficaci nel manifestare differenti tipi di enfasi: negli slogan di Salvini è stato riscontrato un impiego del punto esclamativo in atti espressivi, mentre nei messaggi elettorali di Giorgia Meloni si rileva un uso in atti iussivi o direttivi, i quali esprimono una richiesta, che può variare dall'ordine all'invito (Lala 2018 e 2019). Come evidente, in questa testualità non è veicolato alcun valore semantico di coercizione, ma un incitamento, che va nella direzione del coinvolgimento con l'elettorato (espresso anche nel manifesto in Fig. 42). Si noti, per contro, una diminuzione della funzione d'appello, indiziata dalla comparsa di una isolata forma impersonale, *si vota*, rispetto alle più frequenti *vota* o *voto*, che saranno commentate più avanti: con ogni probabilità, ciò si spiega con il fatto che il manifesto è stato pubblicato il giorno stesso delle elezioni e pertanto acquisisce tratti tipici dell'esercizio della funzione referenziale, dando informazioni sull'orario del voto.

³⁷ Twitter e Facebook Meloni, 23.09.22.

³⁸ Twitter e Facebook Meloni 25.09.22.

³⁹ Par. 3.1.

L'esortativo inclusivo ricompare in altri manifesti, pubblicizzati sulle pagine "social", tanto di Giorgia Meloni, quanto di FdI (e di altri esponenti del partito), in cui l'invito è rivolto simultaneamente all'elettorato e alla propria parte politica, con l'obiettivo di realizzare un contrasto con l'altra parte politica, rappresentata linguisticamente come schieramento politico (*la sinistra*) (Fig. 46) o con un semplice deittico (*li in teniamoli*) (Fig.47)::

Fig. 46⁴⁰Fig. 47⁴¹

Si può notare come l'aperta contrapposizione con la parte antagonista ("loro"/la III persona) sia realizzata, oltre che dai testi, attraverso diversi espedienti: nel primo manifesto (Fig. 46) le proposte politiche del PD vengono poste su un piano subordinato rispetto a quello in cui compare l'immagine della Meloni con il suo messaggio elettorale, rendendole come nascoste, semi-cancellate (ma tutte chiaramente leggibili), vale a dire "neutralizzate" dall'eventuale vittoria di FdI. Da notare che il simbolo del PD appare attraversato da una linea spezzata, rendendo il messaggio, nella sua interezza, non particolarmente pacato (questo espediente grafico è presente anche in manifesti di Salvini in cui viene usato l'*argumentum ad personam*). Nell'altro (Fig. 47), la contrapposizione è ancora più violenta, nel mettere in opposizione non proposte o orientamenti politici, ma persone (qui si riconosce un tratto di personalizzazione che avvicina ancora

⁴⁰ Facebook e Twitter Meloni, 2.08.22.

⁴¹ Facebook Fratelli d'Italia, 29.07.22.

Meloni a Salvini), attraverso espedienti esclusivamente visivi: le immagini dei rappresentanti della sinistra sono in bianco e nero, con espressioni del volto poco rassicuranti, mentre Giorgia Meloni è a colori e sorridente.

Ma, come anticipato, Giorgia Meloni sfrutta un'ampia gamma di mezzi linguistici: in altri manifesti la contrapposizione viene esplicitata proprio attraverso l'uso dei pronomi *Noi* e *Loro*. In particolare, nel manifesto in Fig. 49 emergono tutte e tre le persone: la prima plurale (esclusiva) *noi*, la terza plurale per la controparte, *loro*, e la seconda singolare nell'imperativo *scegli*. Peraltro, si fa notare come questa contrapposizione sia perfettamente parallela a quella di Salvini, riproducendo la scelta di fondo dei due leader: Matteo Salvini privilegia l'“io”, a cui contrappone un “lui/lei” (si vedano le Figg. 17-18) mentre Giorgia Meloni, insiste sul “noi” rispetto al quale crea contrasto con un “loro”. Si fa notare anche in questo caso il ruolo giocato dai simboli di partito: i due manifesti rendono visibile l'evoluzione del simbolo di FdI, che nel primo manifesto non presenta ancora il nome di Giorgia Meloni; nel contempo si rileva nuovamente l'intervento sul simbolo del PD, spezzato da una linea che ne evoca la sconfitta.

Fig. 48⁴²Fig. 49⁴³

⁴² Facebook Meloni, 24.07.22.

⁴³ Twitter Meloni, 19.08.22.

In pochi dei manifesti elettorali di Giorgia Meloni (presenti sui canali “social” Twitter e Facebook, tanto della stessa Meloni quanto di FdI) risulta linguisticamente in primo piano la seconda persona singolare in funzione di appello, prototipicamente rappresentata dalla forma imperativa del verbo, *vota*, *leggi*:

[simboli di FdI e PD con percentuali dei sondaggi] - Il 25 settembre - vota da Patriota

Elezioni politiche - 25 settembre 2022 - [simbolo FdI] - Leggi - il programma - elettorale - dei patrioti

Il 25 settembre, - il futuro dell'Italia - lo deciderai TU. [simbolo FdI] - Vota Giorgia Meloni, - di lei ti puoi fidare

IL 25 SETTEMBRE - SCEGLI LA COERENZA - [simbolo FdI] - VOTA - GIORGIA MELONI.

Proprio l'uso della II persona singolare ha una presa emotiva maggiore rispetto alla corrispondente plurale, dal momento che sembra chiamare in causa specificamente ogni singolo destinatario del messaggio. Non a caso, quest'uso, sebbene solo in parte, è sovrapponibile allo slogan *Scegli*, utilizzato dal PD, forma che compare anche nell'ultima tra le testualità sopra riportate.

L'enfasi sul destinatario in quanto individuo, con evidenziazione grafica del pronome di II singolare, *TU* (reso con caratteri maiuscoli) è realizzata in una sintassi che presenta dislocazione e ordine marcato OVS (*il futuro dell'Italia lo deciderai TU*), in cui, però, è proprio il soggetto ad essere focalizzato; peraltro, l'enfasi sulla II persona è amplificata dalla ripresa tramite la forma obliqua (*ti*) e si realizza in una formulazione linguistica in cui nel gioco dei riferimenti deittici compare, forse come unica volta, il ricorso al pronome di III persona femminile singolare, *lei* (il manifesto è presente sul profilo Facebook di FdI). Da notare che in questa strutturazione la funzione di appello, e quindi l'invito al voto, è realizzata attraverso un imperativo futuro⁴⁴. In questi messaggi elettorali si può osservare il ricorrere della dimensione patriottica, già emersa nel testo *Avanti, Patrioti. Siamo pronti* (Fig. 35), tipica della comunicazione politica della destra, e strettamente connessa alla retorica esaltazione della dimensione nazionale, già vista in Salvini con il richiamo all'Italia, agli italiani e al tricolore.

⁴⁴ Forma riconosciuta solo da alcuni studiosi, tra cui Dardano e Trifone (1997: 324).

Altri manifesti offrono ulteriori spunti di analisi relativamente al gioco della deissi personale; quello in Fig. 50, pubblicato appena due giorni prima del 25 settembre, suggerisce una forte identificazione, di pensieri e intenti, tra il soggetto politico e l'elettorato, mittente e destinatario: in una strutturazione linguistica in cui *Io amo l'Italia* rappresenta il cuore del messaggio elettorale – in caratteri nettamente più grandi rispetto al resto del testo – ed è presentato come l'enunciato di ciascuno degli elettori (*Quando domenica vi recherete a votare, pensate con orgoglio: "Io amo l'Italia"*), si noti come questo sia visivamente associato alla figura di Giorgia Meloni ed è, pertanto, ad un primo sguardo, percepito come enunciazione della leader politica di Fdi.



Fig. 50⁴⁵

Questa strategia, che mira all'inclusività in un modo più sottile rispetto al consueto uso del "noi", si realizza in altri due manifesti, comparsi sui canali social di Fdi, in cui l'identificazione tra mittente e destinatario è veicolata in modo esplicito con la forma verbale *voto*, rafforzata nel primo manifesto dall'enfatico *io*:

IL 25 SETTEMBRE - [simbolo Fdi] - IO VOTO GIORGIA

[simbolo Fdi] - IL 25 SETTEMBRE - VOTO GIORGIA MELONI - SENZA ALCUN DUBBIO.

⁴⁵ Twitter e Facebook Meloni 23.09.22.

Questa tecnica è stata, d'altra parte, già rilevata anche nell'hashtag usato da Salvini #25settebrevotolega e, in misura ridottissima, anche da parte di Letta, #VotoPD e #25settebrevotoPD.

Sempre nel primo manifesto è da notare anche l'uso del nome proprio, che, come è stato rilevato da più parti, rappresenta un ulteriore tratto di personalizzazione, tendenza al populismo e predilezione per un tipo di linguaggio familiare. In questo senso, l'uso del solo nome di battesimo - inaugurato con riferimenti a Silvio, Matteo, Beppe - è stato visto come ulteriore indizio del passaggio al paradigma del "rispecchiamento" (Coletti 2013). D'altra parte, anche la recente campagna elettorale per il Presidente della repubblica francese ha visto l'affermarsi di questo modello con manifesti elettorali di Marine Le Pen, in cui è comparso il solo nome di battesimo della candidata, *Marine Présidente - Femme d'État*, con l'obiettivo di realizzare un avvicinamento con l'elettorato (si veda La Fauci 2022).

Inoltre, la formula elettorale *Io voto Giorgia* evoca immediatamente lo slogan che è stato parte della comunicazione politica di Giorgia Meloni prima della campagna elettorale stessa: *Io sono Giorgia*. In quest'ultimo, la coincidenza tra mittente e destinatario era assente a favore di una forte focalizzazione sul mittente e un prevalere della funzione emotiva, configurando un forte esempio di personalizzazione⁴⁶. È quindi tanto più significativo che Giorgia Meloni abbandoni questa prospettiva comunicativa in occasione della campagna elettorale, in favore di formulazioni linguistiche che, come evidenziato fin qui, sono piuttosto incentrate sulla pluralità inclusiva. A questo proposito, si può notare come i sostenitori della leader di FdI rispondano al gioco di identificazione promosso dagli slogan elettorali, creando cartelloni che riprendono, variandolo, proprio il noto slogan: *NOI SIAMO Giorgia* (Fig. 51), aprendo ad una interpretazione che sta per 'noi siamo te', 'noi siamo tutte le cose che rappresentano la tua identità politica', 'noi siamo una cosa sola'.

⁴⁶ Lo slogan, come noto, è diventato anche il titolo di un libro a firma di Giorgia Meloni (*Io sono Giorgia. Le mie radici, le mie idee*, Milano, Rizzoli, 2021) e ha dato via ad una serie di meme, e di riprese da parte di altri noti personaggi pubblici, tra cui la cantante Giorgia.



Fig. 51



Fig. 52

La rilevanza e diffusione di questo slogan si conferma anche nella ripresa inter-testuale da parte di Emma Bonino che ne riproduce lo schema (in *Io sono EMMA*) con un chiaro intento di contrapposizione, per quanto non esplicita (Fig. 52). In questo senso, la formula *Io sono Giorgia* mostra caratteri propri dello slogan in senso stretto.

4. Conclusioni

L'analisi dei manifesti elettorali prodotti per le politiche del 25 settembre 2022 ha confermato le ipotesi avanzate nell'introduzione. Un primo obiettivo è consistito nel mostrare come le testualità presenti nei manifesti elettorali possano essere considerate forme condensate del discorso politico contemporaneo, manifestandone alcuni caratteri rilevanti, tra cui la personalizzazione e il leaderismo. È stato infatti possibile mettere in evidenza come questo aspetto, tradizionalmente più marcato nella comunicazione politica delle destre, sia fortemente caratterizzante dei manifesti di propaganda elettorale della Lega e di Fratelli d'Italia, mentre si presenti in misura minore nella campagna elettorale del Partito Democratico, che, non a caso, ha aderito alla scelta dominante di far comparire l'immagine del leader solo nella seconda fase della campagna elettorale.

L'indagine ha mostrato come i testi presenti nei manifesti elettorali appaiano tendenzialmente poco raffinati in termini sintattici e retorici e ciò ha portato all'ipotesi di classificarli come "slogan deboli". Rispetto agli slogan in senso stretto, molti dei testi analizzati mancano di alcune caratteristiche

canoniche di questo genere di testualità: impiego di meccanismi di natura eufonica, che garantiscono la memorizzazione e la riproducibilità; uso di figure retoriche; impiego della tecnica della ripetizione, con variazione, di espressioni ampiamente circolanti. Al di là dei pochi casi messi in evidenza (ad esempio, *Ridateci Salvini a difendere i confini; Per la salute, per l'economia, per i nostri figli; Giù le tasse sul lavoro, su i salari; Lei/lui non crede [...] io ci credo; Abbiamo riempito [...] ora riempiamo; io sono Emma*) nei messaggi elettorali analizzati è stato constatato, per lo più, l'impiego di un linguaggio "telegrafico", volto alla semplice evocazione dei temi cardine dello schieramento politico, per quanto tra di loro molto distanti (*Stop Fornero! Sì quota 41; Lavoro sottopagato / Salario minimo*). Se, dunque, il carattere della brevità avvicina il linguaggio dei manifesti elettorali a quello pubblicitario, è possibile sostenere che esso se ne allontani per altri aspetti di rilievo: per questo motivo, in questo lavoro, si è preferito parlare spesso di "testualità", "messaggi elettorali" piuttosto che di "slogan".

L'attenzione è stata dedicata quindi a rilevare e descrivere, nei manifesti oggetto d'analisi, le diverse scelte operate dalle forze politiche relativamente ai meccanismi di riferimento personale, che implicano strategie di avvicinamento e distanziamento tra mittente/politico (I persona), destinatario/elettore (II persona) e terza componente dello schema enunciativo (i fatti, l'altra parte politica, la III persona). Se ai testi presenti nei manifesti elettorali non è assegnato il compito di veicolare contenuti significativi, e tanto meno di strutturarli linguisticamente in un modo che risulti incisivo, appare evidente come essi siano concepiti per realizzare il semplice ruolo di mantenere il dialogo con l'elettorato, esercitando allo stesso tempo funzione fatica e conativa. L'analisi presente in questo lavoro ha mostrato come, nel realizzare questo obiettivo, le diverse forze politiche manifestino strategie differenti, che fanno emergere di volta in volta un ruolo maggiore assegnato al locutore, o al destinatario o alla persona politica del "noi", che può unire locutore e destinatario. L'analisi ha confermato come le scelte delle forze politiche in esame siano polarizzate anche rispetto a questo parametro: la Lega, con Salvini, mostra una dominanza dell'"io" (*Credo*), mentre il PD oscilla tra una III persona (*vincono le idee*) e una II (*Scegli*), laddove Fratelli d'Italia, con Giorgia Meloni, mostra di sfruttare appieno tutte le diverse possibilità, puntando

molto sul “noi”, tanto inclusivo (*teniamo la sinistra lontana; scriviamo la storia insieme; riempiamo le urne*) quanto esclusivo (*Noi siamo pronti; gli italiani sono con noi*), o su una I persona (*io sono pronta*) che manifesta sovrapposizione e identificazione con il destinatario/elettore (*io amo l'Italia, io voto Giorgia*).

In conclusione, anche focalizzando il punto di vista sull'analisi dei riferimenti deittici di persona, è possibile cogliere un forte divario tra le scelte comunicative delle forze contrapposte: la destra, più incline alla personalizzazione, sia nel senso di una identificazione del progetto politico con la persona del leader, sia in quello della rappresentazione di una soggettività condivisa di istanze e sentimenti tra elettori e eligendi, e la sinistra, tradizionalmente volta alla prospettiva referenziale e/o all'appello all'elettore.

Bibliografia

- Antonelli, Giuseppe, 2000, “Sull'italiano dei politici nella Seconda Repubblica”, in Vanvolsem S., Vermandere D., Musarra F. & D'Hulst Y. (a cura di), *L'italiano oltre frontiera*, Firenze, Cesati, vol. 1, 211-234.
- Amenta, Luisa, 2011, “Il linguaggio della politica nella rete”, in Nesi A. (a cura di), *Storia della lingua e storia dell'Italia Unita*, Firenze, Cesati, 87-101.
- Aprile, Marcello, 2021, *Manuale di base di linguistica e grammatica italiana*, Bologna, Il Mulino.
- Bhat, Shankara D.N., 2004, *Pronouns*, New York, Oxford University Press.
- Benveniste, Émile, 1966a, “La nature des pronoms”, in *Problèmes de linguistique générale*, Paris, Gallimard, 1, 251-257.
- Benveniste, Émile, 1966b, “Structure des relations de personne dans le verbe”, in *Problèmes de linguistique générale*, Paris, Gallimard, 1, 225-236.
- Bruno, Marcello Walter, 1996, *Promocrazia: tecniche pubblicitarie della comunicazione politica da Lenin a Berlusconi*, Genova, Costa & Nolan.
- Calabrese, Omar, 1998, *Come la boxe. Lo spettacolo della politica in TV*, Bari, Laterza.
- Campus, Donatella, 2004, “La comunicazione politica di Berlusconi. Percorsi di lettura”, *Comunicazione Politica*, vol. V (2), 179-189.
- Coletti, Vittorio, 2013, *L'italiano della politica*, Accademia della Crusca, <https://accademidellacrusca.it/it/contenuti/l-italiano-della-politica/7372>.
- Corbett, Greville G., 1991, *Gender*, New York, Cambridge University Press.
- Cortelazzo, Mauro, 2017, “Sulla cattiva strada: la lingua politica e l'iperrispecchiamento”, in *Turpiloquio, menzogna, manipolazione. La parola dei politici oggi*, Speciale

- di *Lingua Italiana*, Istituto della Enciclopedia Italiana, Treccani, https://www.treccani.it/magazine/lingua_italiana/speciali/politici/Cortelazzo.html.
- Cosenza, Giovanna, 2012, *SpotPolitik. Perché la «casta» non sa comunicare*, Bari-Roma, Laterza.
- Cosenza, Giovanna, 2018, *Semiotica e comunicazione politica*, Bari-Roma, Laterza.
- Cysouw, Michael, 2013, "Inclusive/exclusive distinction in independent pronouns", in Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*, chapter 39. Leipzig, Max Planck Institute for Evolutionary Anthropology, <https://wals.info/chapter/39>.
- Daniel, Michael, 2013, "Plurality in Independent Personal Pronouns", in Dryer, M. S. & Haspelmath, M. (eds.), *The world atlas of language structures online*, Leipzig, Max Planck Institute for Evolutionary Anthropology, chapter 35, <https://wals.info/chapter/35>.
- Daniel, Michael, 2020, "Person in Morphology", *Oxford Research Encyclopedia of Linguistics*, Oxford University Press, <https://oxfordre.com/linguistics/display/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-544>.
- Dardano, Maurizio & Trifone, Pietro, 1997, *La nuova grammatica della lingua italiana*, Bologna, Zanichelli.
- De Santis, Cristiana, 2016, "Pensiamo, pensavamo e penseremo: strategie di costruzione dell'autorità nel discorso dei nuovi leader", in Librandi R. & Piro R. (a cura di), *L'italiano della politica e la politica per l'italiano*, Firenze, Cesati, 321-332.
- De Santis, Cristiana, 2019, "Fortuna e circolazione discorsiva di alcuni slogan italiani", *Kwartalnik Neofilologiczny* 66 (2), 361-370.
- De Santis Cristiana & Simonini Jessy, 2017, "Renzi: la forza del «noi»", in *Turpiloquio, menzogna, manipolazione. La parola dei politici oggi*, Speciale di *Lingua Italiana*, Istituto della Enciclopedia Italiana, Treccani, https://www.treccani.it/magazine/lingua_italiana/speciali/politici/De_Santis.html.
- Dell'Anna, Maria Vittoria, 2017, "Berlusconi: io, la gente e me", in *Turpiloquio, menzogna, manipolazione. La parola dei politici oggi*, Speciale di *Lingua Italiana*, Istituto della Enciclopedia Italiana, Treccani, https://www.treccani.it/magazine/lingua_italiana/speciali/politici/Dell_Anna.html.
- Dell'Anna, Maria Vittoria & Lala, Pierpaolo, 2004, *Mi consenta un girotondo. Lingua e lessico nella Seconda Repubblica*, Galatina, Congedo.
- Desideri, Paola, 1984, *Teoria e prassi del discorso politico. Strategie persuasive e percorsi comunicativi*, Roma, Bulzoni.
- Desideri, Paola, 1999, "La comunicazione politica", in Gensini S. (a cura di), *Manuale della comunicazione*, Roma, Carocci, 391-418.

- Desideri, Paola, 2011, "Linguaggio della politica", in *Enciclopedia dell'italiano Treccani*, Roma, Istituto della Enciclopedia Italiana Treccani, https://www.treccani.it/enciclopedia/linguaggio-della-politica_%28Enciclopedia-dell%27Italiano%29/.
- Dobrushina, Nina & Goussev, Valentin, 2005, "Inclusive imperative", in Filimonova E. (ed.), *Clusivity: typology and case studies of the inclusive-exclusive distinction*, Amsterdam/Philadelphia, John Benjamins, 179-211.
- Gualdo, Riccardo & Dell'Anna, Maria Vittoria, 2004, *La faconda Repubblica. La lingua della politica in Italia (1992-2004)*, San Cesario di Lecce, Manni.
- Hymes, Dell H., 1974, *Foundations in Sociolinguistics. An Ethnographic Approach*, London, Tavistock.
- Huddleston, Rodney D., & Pullum, Geoffrey K., 2002, *The Cambridge Grammar of the English Language*, Cambridge, Cambridge University Press.
- Jakobson, Roman, 1957, *Shifters, verbal categories and the Russian verb*, Harvard University, Department of Slavic Languages and Literatures, Russian Language Project.
- Jespersen, Otto, 1924, *The philosophy of grammar*, London, George Allen and Unwin.
- La Fauci, Nunzio, 2016, "Noi, persona politica", in Librandi R. & Piro R. (a cura di), *L'italiano della politica e la politica per l'italiano*, Firenze, Cesati, 387-400.
- La Fauci, Nunzio, 2022, "Una contesa di persona, genere, preposizioni, nomi e pronomi. Da 'Emmanuel Macron avec vous' a 'femme d'état' di Marine Le Pen. Analisi della campagna elettorale", *MicroMega*, 8 aprile 2022, <https://www.micromega.net/elezioni-francia-campagna-elettorale/>.
- Lala, Letizia, 2018, "Punto esclamativo", in Ferrari A., Lala L., Longo F., Pecorari F., Rosi B. & Stojmenova R. (a cura di), *La punteggiatura italiana contemporanea. Un'analisi comunicativo-testuale*, Roma, Carocci, 201-215.
- Lala, Letizia, 2019, "Il punto esclamativo e il punto interrogativo nella narrativa contemporanea", in Ferrari A., Lala L., Pecorari F. & Stojmenova Weber R. (a cura di), *Punteggiatura, sintassi, testualità nella varietà dei testi italiani contemporanei*, Firenze, Cesati, 379-394.
- Lombardi Vallauri, Edoardo, 2015, "Pesare l'implicito", in Ferrari A., Lala L. & Stojmenova R. (a cura di), *Testualità. Fondamenti, unità, relazioni*, Firenze, Cesati, 61-81.
- Lombardi Vallauri, Edoardo, 2019, "Sfruttamento di "immagini" implicite nella pubblicità e nella propaganda politica italiana", in Moscarda Mirković E. & Habrle T. (a cura di), *Sguardo sull'immaginario italiano. Aspetti linguistici, letterari e culturali*, Pula, Sveučilište Jurja Dobrile u Puli, 267-294.
- Lombardi Vallauri, Edoardo & Masia, Viviana, 2016, "Misurare l'informazione implicita nella propaganda politica italiana", in Librandi R. & Piro R. (a cura di), *L'italiano della politica e la politica per l'italiano*, Firenze, Cesati, 539-557.

- Lyons, John, 1977, *Semantics*, Cambridge, Cambridge University Press.
- Manetti, Giovanni, 2015, “Il noi tra enunciazione, indessicalità e funzionalismo”, in Janner M.C., Della Costanza M.A. & Sutermeister S. (eds.), *Noi, Nous, Nosotros*. Studi romanzi/Études romanes/Estudios románicos, Bern, Peter Lang, 23-44.
- Ondelli, Stefano, 2017, “Salvini contro tutti: «Preferisco i populistici ai fessi»”, in *Turpiloquio, menzogna, manipolazione. La parola dei politici oggi*, Speciale di *Lingua Italiana*, Istituto della Enciclopedia Italiana, Treccani, https://www.treccani.it/magazine/lingua_italiana/speciali/politici/Ondelli.html.
- Ottaviano, Chiara 2004, “Ricorrenze e novità nella comunicazione politica. Manifesti e slogan nelle campagne elettorali all'alba del nuovo secolo”, in Ridolfi, M. (a cura di), *Propaganda e comunicazione politica. Storia e trasformazioni nell'età contemporanea*, Milano, Mondadori, 287-306.
- Renzi, Lorenzo, 2001, “La deissi. La deissi personale e il suo uso sociale”, in Renzi L., Salvi G. & Cardinaletti A. (a cura di), *Grande grammatica italiana di consultazione*, vol. III, *Tipi di frasi, deissi, formazione delle parole*, 350-375, Bologna, Il Mulino.
- Santulli, Francesca, 2005, *Le parole del potere, il potere delle parole. Retorica e discorso politico*, Milano, Franco Angeli.
- Sergio, Giuseppe, 2008, “La politica al muro: manifesti elettorali e slogan”, in Vetrugno R., De Santis C., Panziera, C. & Della Corte F. (a cura di), *L'italiano al voto*, Firenze, Accademia della Crusca, 5-51.
- Serianni, Luca, 1988, *Grammatica italiana. Italiano comune e lingua letteraria. Suoni, forme, costrutti*, con la collaborazione di A. Castelvechi, Torino, UTET.
- Siewierska, Anna, 2013, “Gender distinctions in independent personal pronouns”, in Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online*, chapter 44. Leipzig, Max Planck Institute for Evolutionary Anthropology, <https://wals.info/chapter/44>.
- Simone, Raffaele, 2013, *Nuovi fondamenti di linguistica*, Milano, McGraw-Hill.
- Spina, Stefania, 2012, *Openpolitica. Il discorso dei politici italiani nell'era di Twitter*, Milano, Franco Angeli.
- Uspenskij, Boris, 2008, “Deissi e comunicazione. La realtà virtuale del linguaggio”, in Keidan A. & Alfieri L. (a cura di), *Deissi, riferimento, metafora: questioni classiche di linguistica e filosofia del linguaggio*, Firenze, Firenze University Press, 107-63.
- Vincent, Nigel, 1980, “Words versus Morphemes in Morphological Change: The Case of Italian -amo”, in Fisiak J. (ed.), *Historical Morphology*, The Hague, De Gruyter Mouton, 383-398.
- Viviani, Andrea, 2011, “Slogan”, in *Enciclopedia dell'Italiano*, Treccani, Istituto della Enciclopedia Italiana, https://www.treccani.it/enciclopedia/slogan_%28Enciclopedia-dell%27Italiano%29/.

FRANCO LORENZI*

Parlare di *niente*. Appunti per un lessico dell'inesistente

Abstract

Negli studi linguistici contemporanei sono stati rivisitati alcuni temi classici della speculazione filosofica, come l'*essere* e la *definizione*, che hanno attirato l'interesse dei linguisti per una diversa prospettiva d'indagine. In questo articolo proponiamo di allargare l'interesse a un tema altrettanto affascinante: *niente*. *Niente* e molti altri lessemi (a partire dal quasi-sinonimo *nulla*) formano il *lessico dell'inesistente*, presente in italiano come in tutte le lingue naturali. Per iniziare il lavoro, discuteremo alcune descrizioni metalinguistiche.

Parole chiave: lessicologia, semantica, metalinguaggio

In contemporary linguistic studies, some classical topics of philosophical speculation, such as *being* and *definition*, have been revisited, attracting the interest of linguists for a different perspective of inquiry. In this article, we propose to extend this interest to an equally fascinating topic: *niente* (*nothing, nothingness*). *Niente* and many other lexemes (starting with the near-synonym *nulla*) form the lexicon of the non-existent, present in Italian as in all natural languages. To begin the work, we will discuss some metalinguistic descriptions.

Keywords: lexicology, semantics, metalanguage

Introduzione

Negli studi di linguistica contemporanea sono stati rivisitati alcuni temi classici della speculazione filosofica, che hanno attirato l'interesse dei linguisti per una diversa prospettiva d'indagine. Il riferimento più immediato è all'*essere* e alla *definizione*, oggetto di un'attenzione millenaria a partire dalla filosofia greca antica.

* Franco Lorenzi, Università degli Studi di Perugia, franco.lorenzi@unipg.it.

In questo breve articolo proponiamo di allargare l'interesse a un tema altrettanto affascinante: *niente*. Il lavoro che presentiamo vuole avviare una riflessione sistematica su *niente* e i molti lessemi (a partire dal quasi-sinonimo *nulla*) che formano il *lessico dell'inesistente*, presente in italiano come in tutte le lingue naturali. Effettueremo una prima ricognizione sulle descrizioni metalinguistiche legate a *niente* nella semantica strutturale e cognitiva, nelle opere di Robert Martin, George A. Miller e Anna Wierzbicka indicando alcune linee di ricerca che ci sembrano particolarmente promettenti. Da questo, il carattere di 'appunti' con il proposito di avviare una ricerca sistematica.

1. Come ben sappiamo, nella cultura occidentale la speculazione filosofica e la riflessione sul linguaggio appaiono inestricabilmente collegate. Partendo dalla filosofia greca, un riferimento immediato è all'*essere* e alla *definizione*. L'*essere* costituisce un tema fondante e ad esso si collegano l'*essenza*, l'*esistenza* e la *categorizzazione*. La costruzione di un sistema di categorie del pensiero si basa sulla concezione di *ciò che è*; come afferma Benveniste (1958/1966: 70) discutendo il sistema aristotelico:

Au-delà des termes aristotéliennes, au-dessus de cette catégorisation, se déploie la notion d' « être » qui enveloppe tout. Sans être un predicat lui-même, l' « être » est la condition de tous les prédicats. Toutes les variétés de l' « être-tel », de l' « état », toutes les vues possible du « temps », etc., dépendent de la notion d' « être ».

Ma Benveniste aggiunge che la possibilità di sviluppare “toute la métaphysique grecque de l' « être »” si è basata sulla particolare struttura del greco, che metteva a disposizione un elemento flessibile e multiforme, mancante, ad esempio, nell'Ewe. L' “εἶναι” consentiva di “faire de l' « être » une notion objectivable, que la réflexion philosophique pouvait manier, analyser, situer comme n'importe quel autre concept” (p. 71). In questo senso, la lingua si conferma come la struttura che “donne sa *forme* au contenu de pensée”; in modo esplicito, Benveniste rileva che “[l]a forme linguistique est donc non seulement la condition de transmissibilité, mais d'abord la condition de réalisation de la pensée”. (p. 64)

Ancora, ripercorrendo la storia del verbo *essere*, Moro (2010) mette in evidenza aspetti che gettano una luce nuova sulla questione: la varietà delle fenomenologie lessicali delle diverse lingue, la dimensione combinatoria e

pragmatica (temi su cui torneremo più avanti), per giungere alla conclusione che “[i]l verbo *essere*, dunque, è vuoto, come sono vuoti tutti gli elementi che compongono la grammatica”. *Essere* offre, in certe lingue, ma non in tutte, la possibilità di segnalare la proiezione che la lingua opera “sulla realtà del mondo per il tramite della grammatica” (p. 240), in particolare per l’interpretazione del tempo fisico. “Il tempo della grammatica, infatti, non è un semplice riflesso del tempo fisico, ma costituisce una struttura interpretativa specifica di questo «fenomeno»”; il tempo grammaticale, allora, è l’elemento universale (insieme linguistico e cognitivo), mentre *essere* è uno degli strumenti che, insieme ad altri, può essere usato nelle lingue per esprimerlo.

La definizione (Peruzzi 1983, Sager 2000) è lo strumento euristico che consente di determinare caratteristiche e proprietà di un *definiendum* *x* usando un *definiens* *y*. La relazione tra *x* e *y* dà la possibilità, in questo modo, di chiarire le relazioni tra parole e cose e tra parole tra di loro (*word-thing* e *word-word definitions* nella terminologia di Robinson 1954). In ambito linguistico, Bierwisch & Kiefer (1969) hanno studiato la creatività lessicale delle definizioni, distinguendo tra definizioni *analitiche* e *introduttive*. Le prime stabiliscono un’equivalenza lessicale tra il *definiendum* e il *definiens* e, quindi, non aggiungono nuove parole nella lingua, mentre le seconde ampliano il patrimonio lessicale. La distinzione recupera il valore discorsivo delle definizioni e l’alternanza tra *tema* e *rema*. Se le espressioni *x* e *y* sono note sia al mittente che al ricevente, la definizione non porta ad un arricchimento lessicale, ma segnala soltanto la possibilità di usare *x* al posto di *y* nei diversi contesti. Al contrario, se tra *x* e *y* c’è un’asimmetria informativa, la definizione porta alla creazione di una nuova parola (come espressione fonica e come contenuto concettuale), arricchendo la competenza lessicale.

Proseguendo su questa linea, Robert Martin ha considerato il valore *naturale* della definizione. In un noto articolo dedicato alla *définition naturelle* così si esprime:

la *définition naturelle* vise à saisir le contenu naturel des mots, c’est-à-dire le contenu plus ou moins vague que spontanément - et souvent inconsciemment - les locuteurs *y* associent. La *définition naturelle* est ainsi plus ou moins juste. Son contenu évolue avec celui des objets qu’elle entend cerner (Martin, 1990: 87).

2. Veniamo adesso al tema centrale del nostro lavoro: *niente* e il lessico dell’inesistente. L’affermazione che costituisce il titolo di questo articolo (e che ri-

corre in molti studi di carattere filosofico) è, in realtà, una domanda costante in filosofia:

Peut-on penser et dire le rien? Ou bien, parler du rien, est-se déjà ne rien dire – c'est-à-dire se contre-dire? Depuis l'interdit promulgué par Parménide, la philosophie n'a cessé d'explorer les deux branches de cette alternative: soit pour nier le bien-fondé de toute pensée du rien, soit pour tenter d'en tracer les limites légitimes (Claude Romano in Laurent & Romano 2011b: 11)

Parlare di *niente* sembra costantemente un paradosso, legato alle dicotomie *essere/non-essere* e *possibile/impossibile* e alla *necessità*. Ma il problema attraversa tutta la storia della filosofia (Givone 2003), da Parmenide ad oggi, testimoniato da pagine fondamentali come la tavola del *niente* che conclude l'*Appendice* del secondo libro dell'*Analitica trascendentale* della *Critica della ragion pura*¹.

Sintetizzando “[l]e molteplici elaborazioni del concetto di nulla (o *niente*) che sono state prodotte lungo la storia della filosofia occidentale”, Fanciullacci (2006: 7697-7698) individua due concezioni generali: “la prima è quella del nulla come non-essere, ossia come assoluta assenza di ogni positività: il *nulla assoluto*. [...] la seconda [...] è quella per cui esso è la negazione di un dominio determinato dell'essere: è il *nulla relativo*”.

Laurent (2006b: 8-9), più analiticamente, distingue cinque sensi nel “dire le néant”, passando dal “néant suessentiel de l'Un ou de Dieu” al “non-être de l'alterité”, al “non-étant du devenir”, al “presque-rien de la matière” fino al “rien radical”. Entra in gioco la negazione: “[l]a negazione, e quindi il linguaggio”, come afferma Cimatti (2013: 42) discutendo l'opera di Wittgenstein, “trasforma il «mondo» in «realtà»”: “Pensare il mondo com'è significa anche pensare il mondo come *non è*”. E questo è possibile soltanto con il linguaggio; con le parole di un linguista, “la negazione pertiene solamente al livello della pura rappresentazione linguistica *del* mondo, non è *nel* mondo” (Moro 2010: 172)

Infine il *nulla* si collega alla *libertà*. Come afferma Givone (2003: 55) parlando dell'*Apocalisse*, “Il nulla che, negando la ragione di ciò che è (perché

¹ La traduzione della *Kritik der reinen Vernunft* kantiana a cui facciamo riferimento è quella di G. Gentile e G. Lombardo Radice con la revisione di V. Mathieu (1959), condotta sulla seconda ed. del 1787.

questo, perché quello, perché in generale qualcosa e non il nulla), restituisce l'essere a se stesso. Alla sua estasi, al suo estatico stare. Dunque alla libertà”.

3. Insieme al *niente* e al *nulla*, la speculazione filosofica ha lungamente trattato gli *oggetti inesistenti* e, com'è noto, in epoca contemporanea, la discussione su questo tema ha avuto un rinnovato interesse con l'opera di Alexius Meinong (Perszyk 1993).

Antony Everett nel suo libro *The Nonexistent* (2013: 1) elenca “fictional and mythical objects, and sometimes the entities posited by failed scientific theories and the denizens of our dreams and hallucinations”², oggetti che abbiamo presenti nella nostra mente, ma di cui non possiamo dimostrare l'esistenza. Oggetti, però, di cui *parliamo* e che influenzano concretamente la nostra vita, come quando andiamo al cinema o a teatro per vedere materializzata l'esistenza dei pirandelliani *sei personaggi in cerca d'autore*.

Presentando le ricerche moderne e contemporanee, Reicher (2022) pone al centro del problema il rapporto tra *essere* ed *esistere*, tra *c'è* ed *esiste*. Nella tradizione di Frege e Quine non viene rilevata la differenza tra questi usi che, nelle formule logiche, sono resi con l'uso del quantificatore esistenziale \exists che acquista, quindi, “ontological import”: “ $\exists x (...x...)$ ”. Al contrario, nella tradizione meinongiana si distingue tra questi usi, introducendo nelle formule il predicato di esistenza $E!$: “ $\exists x (E!x)$ ”.

La questione ha caratterizzato praticamente tutta la filosofia del Novecento, con apporti fondamentali come la teoria delle descrizioni definite di Russell, e rappresenta tutt'oggi un complesso ambito di indagine per moltissimi ricercatori. In questa sede, citeremo soltanto due autori che esprimono nei loro lavori le due posizioni divergenti. Nello studio già citato, Everett (2013: 3) sostiene che postulare l'esistenza di oggetti inesistenti porti a un'ontologia in cui troviamo “ontically vague objects, cases of vague existence, genuine totalities of indeterminate cardinalities, and to entities which flout the laws of logic and identity”. Più utile, per l'A., è ipotizzare che tutto questo possa avvenire “not in reality, but merely within the scope of the make-believe that there are fictional characters”. Si tratta, quindi, di elaborare una teoria co-

² A questo elenco si possono aggiungere i numeri e molto altro ancora, cfr. Azzouni (2010).

gnitivamente motivata che parta dal rapporto (spesso, l'identità) tra credenze e immaginazione, sul modello degli stati mentali fittizi di Nichols & Stich (2000).

D'altro lato, Terence Parsons (1980) sostiene che un'ontologia in cui ci sia posto per oggetti inesistenti è pienamente legittima e ci consente proprio di *parlare* delle esperienze quotidiane come, ad esempio, i sogni:

The claim here is that it is legitimate to assume that there are physical objects because, among other things, this shared assumption greatly simplifies reports of our experiences. I claim that it is legitimate to assume that there are nonexistent objects for exactly the same reason – it simplifies reports on our experiences in exactly the same way (Parsons 1982: 370)

4. Nel titolo di questo articolo *niente* è in carattere corsivo e non preceduto da un determinativo proprio per indicare il *focus* del nostro progetto: lo studio lessicale delle espressioni che, nelle lingue, testimoniano il modo in cui i parlanti usano parole per l'inesistente. Questo lessico è complesso. Limitiamoci alla lingua italiana e constatiamo subito che esistono due lemmi tra loro ampiamente sinonimi: *niente* e *nulla*. I due lemmi compaiono nel *Vocabolario degli Accademici della Crusca* fin dalla prima edizione (1612: vol. 3, 343); in essa troviamo che *Niente* è glossato in senso negativo come (a) *Nulla, Non punto, Voce, che denota privazione, e negazione*, ma anche in senso positivo come (b) *alcuna cosa (Quando si usa per via di domandare, o di ricercare, o anche di dubitare, ha senso affermativo)*. A sua volta, *Nulla* (vol. 3, 367) è glossato con *Niente* e l'avvertenza immediata che *come quella stessa particella interamente si regola*. Naturalmente *niente* e *nulla* sono utilizzati estesamente nelle glosse di altri lemmi e in molte espressioni polirematiche.

Passando alla situazione attuale, il *Nuovo vocabolario di base della lingua italiana*³ (De Mauro 2016), descrive i due lemmi come appartenenti al vocabolario fondamentale. Per *niente* sono registrati quattro usi: (a) pronome indefinito invariabile; (b) sostantivo maschile invariabile; (c) aggettivo invariabile, e (d) avverbio. Nelle glosse abbiamo *nessuna cosa, nulla, l'assenza di qualcosa, il nulla* accanto a *qualcosa, inezia, poca cosa*. Il lemma è presente, inoltre, in una trentina di espressioni polirematiche.

³ Cfr. *Vocabolario di Base on-line* (<https://dizionario.internazionale.it/>).

Per *nulla* sono registrati tre usi: (a) pronomi indefinito invariabile; (b) sostantivo maschile invariabile e (c) avverbio. Nelle glosse abbiamo *nessuna cosa, niente, ciò che non esiste, condizione di inesistenza* accanto a *qualcosa, inezia, poca cosa*. Il lemma è presente, inoltre, in una quindicina di espressioni polirematiche.

Sia *niente* che *nulla* sono legate alla voce *non essere*, come termine specialistico filosofico che indica *la condizione di non aver vita, di non avere realtà*. Per *nulla* è registrato l'uso tecnico in logica formale per indicare *la classe che non ha alcun elemento*.

Il VdB registra per *niente* e *nulla* una varietà d'usi che spazia dalle locuzioni pragmatiche (*non è niente / non è nulla*) al figurato (*essere niente / essere nulla*). L'elenco degli usi e delle espressioni a diversi livelli di idiomatichità si allunga notevolmente consultando i vocabolari contemporanei, e le conversazioni quotidiane testimoniano che *niente* e *nulla* sono ben presenti nella competenza lessicale del parlante.

Al di là della tendenza a fornire una definizione circolare, la glossa lessicografica ci consegna una tensione tra un valore negativo (*niente* e *nulla* come risultato della negazione applicata a un *x* che perde, in questo modo, la sua esistenza; più radicalmente ancora, essi indicano la mancanza totale di esistenza) e un valore positivo (*niente* e *nulla* come qualcosa di esistente, ma comunque di scarso rilievo, in termini di quantità o valore).

5. A partire da *niente* e *nulla* si sviluppa un lessico ampio e variegato. Da derivati e composti (*annientare, annullare, nullità, nonnulla, nientemeno, nullaosta, nullatenente*) si passa ai lessemi che incorporano nel loro significato l'*inesistente*. Così pensiamo ad *annichilire* (e derivati), che segnala il *ridurre a niente* o a *zero* che "indica la mancanza di ogni unità, cioè è il numero cardinale dell'*insieme vuoto* (o privo di elementi)" (Vocabolario Treccani *on-line* <https://www.treccani.it/vocabolario/>). Ancora, è il "punto di annullamento": "uno zero di una funzione $f(x)$ è un qualsiasi valore della x per il quale la funzione si annulla, cioè assume valore uguale a zero; di qui il modo prov. *zero via zero fa* (o *dà*) *zero*, dal niente non si ricava niente".

La definizione di *zero* ci porta a considerare *vuoto, vuotare* e la lunga serie di lessemi che fanno riferimento alla *mancanza*: da *assente* ad *assentarsi*, da

comparire a scomparire, da scarico a esaurito. In tutti questi casi siamo di fronte a una situazione in cui l'esistente diventa inesistente o viceversa (riempire, presentarsi e così via).

Si apre, poi, il vasto campo degli indefiniti, a partire da *nessuno* e *nessuna*, che nelle lingue formano un insieme eterogeneo; come afferma Martin (1966: 11):

les indéfinis surprennent par un ramassis de formes disparates, étrangement hétéroclites, que n'assemble aucun lien de parenté morphologique. Quel rapport entre *tout* et *chaque*, pourtant si proches de sens? entre *rien* et *quoi que ce soit*, entre *plusieurs* et *quelques*? En fait, la notion d'indéfini couvre un champ grammatical riche de nuances et d'aspects divergents qui expliquent la disparate du système sémiologique.

La presenza degli indefiniti nelle lingue naturali (cfr., ad esempio, Dryer 2013 e Haspelmath 2013 in WALS - Dryer & Haspelmath 2013) testimonia in modo evidente la relazione tra quantità e negazione e su questo torneremo in seguito.

6. Nell'ambito del lungo lavoro di realizzazione del *Trésor de la Langue Française* (oggi disponibile in formato elettronico *TLFi : Trésor de la langue Française informatisé*, <http://www.atilf.fr/tlfi>, ATILF - CNRS & Université de Lorraine), Robert Martin ha pubblicato un celebre saggio su *Le mot "rien"* (1966), che costituisce un punto di partenza obbligato per chi si occupa di questo ambito lessicale. Il libro ha una prospettiva sia sincronica che diacronica, basandosi sulla ricca documentazione disponibile per il *Trésor*.

Martin colloca lo studio di *rien* nel quadro del lessico complessivo della lingua, secondo la visione sistematica che caratterizza gli studi strutturalisti. L'uso di un *mot* non è isolato, ma si collega ai *concurrents* e si comprende pienamente attraverso un esame delle caratteristiche combinatorie.

L'A. parte dall'uso di *rien* nelle espressioni negative e, in particolare, dalla combinazione con *ne*. Riprendendo la definizione di Damourette - Pichon (1928) parla di uso *forclusif*: "*Rien comme pas, goutte, personne*, etc. complète la négation introduite par *ne* et forclot de la réalité ce qui n'existe pas (Martin 1966: 21).

Secondo Martin, *rien* chiude il processo di negazione aperto da *ne*, elemento *discordantiel*. "Le discordantiel *ne* signifie qu'une orientation négati-

ve”, ma questo orientamento è aperto e imperfetto e viene concluso dal secondo elemento, che può variare il campo della negazione:

La comparaison de *ne... rien* avec *ne... pas* ou *ne... point* éclaire l'originalité propre de chacun de ces tours. A opposer, en effet, « Je ne l'examine pas » et « Je n'examine rien », on conclut de la disparition du pronom *le* à la fonction objet du forclusif *rien*. Alors que *pas* se borne à transcender la négation du verbe, *rien* nie tout le champ des objets possible et parfait de ce mouvement négatif la négation immanente du discordantiel (p. 23).

In sintesi, la negazione aperta da *ne* ha un carattere immanente, legato al verbo, che diviene *transcendente* quando il processo è concluso da *rien*, che annulla il campo degli oggetti (nominali) possibili. Ma ancora, *rien* mostra un uso dinamico e può collegarsi all'elemento *discordantiel* per proseguire la negazione in modalità immanente (*ne... rien que ...*) o per costituire un elemento autonomo, nominale.

Martin introduce nella descrizione un elemento fondamentale: il *continuum*. Questo elemento è ben noto nella tradizione strutturalista, che costantemente fa riferimento alla *massa del pensiero* che prende forma nelle lingue; così Eco (1975: 77) parla di “un *continuum* di possibilità fisiche, di eventi psichici, di comportamenti e di pensieri a cui il sistema ha conferito un ordine selezionandovi un insieme strutturato di unità semantiche”.

Martin interpreta il *continuum* in un modo complesso (rappresentato graficamente nei numerosi schemi del suo libro), che collega aspetti linguistici e cognitivi. *Rien* svolge il suo *cinématisme négatif* sul piano sintattico e contemporaneamente semantico, passando dall'essere un elemento grammaticale che completa il processo negativo fondato sul verbo alla piena *sostanza* di un elemento nominale:

Alors qu'en situation forclusive et en corrélation avec le discordantiel, *rien* niait un champ de virtualités sans avoir de lui-même aucune substance propre, aucune « masse sémantique » autonome, il acquiert désormais le poids d'un mot « plein », c'est-à-dire qu'il nie en dehors de toute virtualité, par sémantisme propre (p. 109)

Martin mette a contrasto l'uso grammaticale di *rien* in *Il ne voit rien* rispetto all'uso nominale in *Cela vaut mieux que rien*. Nel primo caso, “le forclusif n'avait de lui-même aucun contenu notionel. Sa fonction, proprement grammaticale, consistait à refuser un ensemble de virtualités, en l'occurrence,

tous les objets possible du verb *voir*”. Nel secondo caso, *rien* completa il movimento nella dimensione della *nullitude*: “le morphème *rien* [...] oppose à «Cela vaut mieux », sa signification désormais autonome de *zéro*, de *néant*, de *vide* : et il a le sens de « ce qui n’est rien »”.

La sostanza a cui dà forma *rien* come elemento nominale è “pauvre mais nettement définie et indépendante des virtualités de sa fonction”. La dimensione della *nullitude* è il *continuum* su cui *rien* raggiunge l’estremo:

Lorsque le mot *rien*, en effet, atteint sans équivoque la nullitude absolue, par exemple dans « Dieu créa le monde de *rien* » (de *zéro*), ou encore dans « Cela vaut mieux que *rien* » (que *zéro*), il atteint le stade ultime de son mouvement négatif. De signification autonome, il n’est plus lié à aucune virtualité et fonctionne définitivement par sémantisme propre (p. 111)

La *nullitude* è una sostanza “pauvre, sans doute, mais réelle” (p. 159). Al termine del *continuum*, *rien* è sinonimo di *néant* e *zéro*, ma nei valori intermedi (segnalati in francese dalla combinazione con gli indefiniti *un* o *des*) esso mantiene l’indicazione di un’esistenza positiva, via via più ridotta e può essere sinonimo di “*bagatelle, broutille, babiole, vétille, etc.*” (p. 160).

7. La *Psycholexicology* di Miller & Johnson-Laird (1976) rimane uno dei tentativi più sistematici di descrizione linguistica di un intero lessico e, non a caso, è diventata la base per il progetto più ambizioso degli ultimi decenni: *WordNet*⁴ (<https://wordnet.princeton.edu/>)⁴. In *Language and Perception* gli AA. basano la dettagliata descrizione del lessico inglese sul rapporto tra operazioni percettive e operazioni concettuali. Partendo da un esame dei risultati delle ricerche psicologiche (e non solo) sulla percezione, gli AA. elencano un insieme di operazioni percettive basiche (1976: 113-115) che costituiscono “the predicates denoting attentional-judgmental abstractions that are possible for the perceptual system”. Le operazioni riguardano oggetti, spazio, tempo ed eventi, come anche intenzioni, memorie e sentimenti. Nella mente dei parlanti i predicati percettivi consentono di attivare procedure cognitive, *test*, che permettono al

⁴ V. i saggi raccolti in Miller (1990) e Fellbaum (1998). In questa sede non prenderemo in considerazione i *WordNets* elaborati per lingue diverse dall’inglese, come ad esempio in EuroWordNet (<https://archive.illc.uva.nl/EuroWordNet/>).

parlante di usare un'etichetta linguistica (*labeling routines*). L'uso di parole si attiva se il *test* dà un risultato positivo, altrimenti l'attivazione non avviene:

The existence in English of such general word as “thing” (in the sense applicable to any percept judged to be a three-dimensional object) means that there is *some* label you can apply to every concrete object, even when you are uncertain about the applicability of a more specific label (Miller & Johnson-Laird 1976: 217-218)

Non c'è posto per la negazione e, più in generale, per l'*inesistente*:

We have so far avoided the syntactic and logical complexities of negation, but it should be clear that there is little place for negation in the kind of descriptions we have proposed. The description is a string of pointers to perceptual, memorial, and conceptual entities; in our formulation, we have no way of assigning a pointer to such nonexistent entities as not-percepts, not-memories, or not-concepts. Negation, in procedural terms, means that the outcome of a “test” instruction should be negative. (Miller & Johnson-Laird 1976: 205).

In WordNet questa posizione si modifica. Il *database* lessicale è organizzato in *synsets*, ovvero *cognitive synonyms, each expressing a distinct concept*. L'impostazione è, quindi, di carattere semantico (e terminologico), per cui si cerca di individuare un (ogni) singolo concetto e di porlo in relazione con un insieme di espressioni linguistiche. In WordNet ogni nucleo concettuale è posto in relazione con una pluralità di espressioni linguistiche, che hanno appunto in comune questo nucleo. La relazione semantica pertinente è quella di *sinonimia parziale o locale*; un insieme di sinonimi è scelto sulla base dei contesti in cui si manifesta il nucleo concettuale comune. I *synsets* sono collegati tra loro da relazioni semantiche, in particolare iponimia/iperonimia e meronimia/olonimia. Questo consente di generare una complessa rete concettuale (di migliaia di concetti) in cui si ricostruiscono catene che portano a un nucleo di concetti fondamentali. WordNet è, quindi, un sistema di ontologie e la *top ontology* termina con il *synset* “*entity*”.

In WordNet concorrono a formare i *synsets* parole appartenenti alle classi lessicali *nome, verbo, aggettivo e avverbio* e l'accesso al *database* avviene attraverso una consultazione per parole inglesi che appartengono a una di queste classi. Per ogni parola è poi fornito l'elenco degli usi nelle quattro classi (o in meno classi, se è il caso) e questo consente di recuperare la dimensione della polisemia lessicale, altrimenti persa nell'impostazione monosemica.

Se inseriamo *nothing* il *synset* nominale in cui compare il lessema comprende *nil, nix, nada, null, aught, cipher, goose egg, naught, zero, zilch, zip, zippo* e la glossa “a quantity of no importance”; il *synset* avverbiale non comprende sinonimi e ha la glossa “in no respect; to no degree”. Scorrendo la catena degli iponimi troviamo soltanto quella del *synset* nominale, con due *synsets*: (1) *nihil* (Latin *nil; nothing* “*nihil habet*” (con la glossa “as used by a sheriff after an unsuccessful effort to serve a writ”) e (2) *bugger all, fuck all, Fanny Adams, sweet Fanny Adams* (con la glossa “little or nothing at all”). Nella catena degli iperonimi abbiamo quattro livelli che portano progressivamente alla radice dell’ontologia:

- (1)
 - (1) *relative quantity* (“a quantity relative to some purpose”)
 - (2) *measure, quantity, amount* (“how much there is or how many there are of something that you can quantify”)
 - (3) *abstraction, abstract entity* (“a general concept formed by extracting common features from specific examples”)
 - (4) *entity* (“that which is perceived or known or inferred to have its own distinct existence (living or non living)”)

Secondo la costruzione di WordNet, questa stessa catena di iperonimi è condivisa da tutte le parole collegate nel *synset* (e quindi, ad esempio, da *zero*).

Con *nothingness* abbiamo soltanto due *synsets* nominali: (1) *void, nullity, nihility* (con la glossa “the state of nonexistence”, e (2) *wind, malarkey, malarky, idle words, jazz* (“empty rhetoric or insincere or exaggerated talk”). Scorrendo la catena degli iponimi per il primo *synset* troviamo soltanto *thin air* (“nowhere to be found in a giant void”), mentre per gli iperonimi abbiamo un’interessante catena ontologica:

- (2)
 - (1) *nonexistence, nonentity* (“the state of not existing”)
 - (2) *nonbeing* (“the state of not being”)
 - (3) *state* (“the way something is with respect to its main attributes”)
 - (4) *attribute* (“an abstraction belonging to or characteristic of an entity”)
 - (5) *abstraction, abstract entity* (“a general concept formed by extracting common features from specific examples”)
 - (6) *entity* (“that which is perceived or known or inferred to have its own distinct existence (living or non living)”)

Il secondo *synset* non ha iponimi e mostra, invece, questa lunga catena ontologica, assai diversa dalla precedente:

- (3)
- (1) *talk, talking* (“an exchange of ideas via conversation”)
 - (2) *conversation* (“the use of speech for informal exchange of views or ideas or information etc.”)
 - (3) *speech, speech communication, spoken communication, spoken language, language, voice communication, oral communication* (“language”)
 - (4) *auditory communication* (“communication that relies on hearing”)
 - (5) *communication* (“something that is communicated by or to or between people or groups”)
 - (6) *abstraction, abstract entity* (“a general concept formed by extracting common features from specific examples”)
 - (7) *entity* (“that which is perceived or known or inferred to have its own distinct existence (living or non living)”)

La catena ontologica in (2) è condivisa da *void* e anche da *nonexistence* e *nonentity*, mentre per quanto riguarda gli iponimi, *nonexistence*, *nonentity* e *nonbeing* mostrano una relazione interessante con:

- (a) *unreality, irreality* (“the state of being insubstantial or imaginary; not existing objectively or in fact”) - proseguendo abbiamo la relazione con *falsity, falseness* (“the state of being false or untrue”)
- (b) *impossibility, impossibleness* (“incapability of existing or occurring”) - proseguendo con *inconceivability, inconceivableness* (“the state of being impossibile to conceive”)

A nostro avviso, la formalizzazione di WordNet rivela due aspetti importanti. In primo luogo, la descrizione del lessico dell'inesistente in WordNet ricorre alla costruzione di oggetti astratti (stati e entità) generati dalla negazione. *Not* e *non* nel *database* sono elementi avverbiali senza iponimi o iperonimi, quindi una sorta di concetti primitivi che operano sui *synsets*. Così le concettualizzazioni lessicali di esistenza, realtà, possibilità e verità diventano la base per nuove concettualizzazioni lessicali. In secondo luogo, i lessemi coin-

volti mostrano una polisemia costante che rinvia alle considerazioni fatte in precedenza parlando dell'analisi strutturale di *rien*.

8. Concludiamo il nostro sintetico panorama delle descrizioni metalinguistiche del lessico dell'inesistente con il *Natural Semantic Metalanguage*⁵, una delle teorie cognitive che ha avuto il maggiore sviluppo negli ultimi decenni. A partire dal celebre *Lingua Mentalis* (1980) di Anna Wierzbicka, il NSM vanta centinaia di pubblicazioni e ricerche dettagliate su oltre trenta lingue, tra loro diverse per famiglie e diffusione⁶. Anna Wierzbicka riporta la base filosofica del suo lavoro alle speculazioni di Leibniz ed usa il metodo di analisi delle *parafrasi riduttive*. Nell'analisi lessicale il ricercatore riformula il significato di una parola o di un'espressione complessa usando (a) parole più semplici e (b) in modo tendenzialmente esaustivo, con una lunga definizione che ha un carattere enciclopedico. Questa operazione porterebbe a una circolarità se non fosse interrotta al momento in cui il ricercatore si accorge che esistono parole *indefinibili*, il cui significato è auto-evidente⁷. Queste parole sono dette *semantic primes* e formano il *semantic core* della lingua insieme a strutture sintattiche minimali che le legano.⁸ Secondo quanto troviamo nel sito del NSM:

When Wierzbicka and colleagues claim that DO, BECAUSE and GOOD, for examples, are semantic primes, they are claiming (i) that these words are essential for explicating the meanings of numerous other words and grammatical constructions, and (ii) that they cannot themselves be explicated in a non-circular fashion, i.e. that their meanings are conceptually simple.

I ricercatori hanno operato nella convinzione che i *semantic primes* preliminarmente individuati per l'inglese in realtà abbiano una validità universale e che, quindi, siano presenti in tutte le lingue. Per questo motivo, il NSM

⁵ Per la documentazione sul NSM rinviamo al sito del progetto (<https://intranet.secure.griffith.edu.au/schools-departments/natural-semantic-metalanguage> e al sito <https://nsm-approach.net/>, con informazioni e articoli.

⁶ In realtà, le lingue citate nelle pubblicazioni del NSM sono più numerose, proprio per la volontà di mostrare il valore universale delle ricerche svolte.

⁷ Una strategia identica è usata nella *Functional Grammar* sviluppata da S.Dik (1978).

⁸ Queste strutture combinatorie hanno una forma basica predicato-argomenti.

distingue tra *semantic primes* ed *exponents*, i.e. realizzazioni dei *primes* nelle singole lingue. Combinando i *primes* in strutture più complesse, si otterranno poi le *semantic molecules*, cioè i diversi lessemi presenti nel lessico delle lingue. Sempre nell'intento di cercare universali semantici, Anna Wierzbicka, Cliff Goddard e i ricercatori del NSM hanno individuato un ristretto numero di *molecules* che hanno un valore quasi-universale, poiché si trovano praticamente in tutte le lingue.

L'insieme dei *primes*, delle *molecules* e delle strutture predicative di base formano quello che viene definito *Basic Human*, "a minimal language which corresponds by and large to the shared lexical, grammatical and conceptual core of all (or nearly all) languages" (Wierzbicka 2018: 22). Come abbiamo accennato (e come ripete costantemente Anna Wierzbicka), il fondamento filosofico del NSM risiede nella concezione leibniziana dell'*alphabetum cogitationum humanarum* "an alphabet of human thoughts" (Wierzbicka 2018: 27), ovvero il corredo concettuale umano innato. Usando le parole e le costruzioni del *Basic Human* (realizzate dai diversi *exponents* nelle diverse lingue) otteniamo anche il vantaggio di poter *parafrasare per tradurre*, cioè rendere le espressioni complesse di una lingua in un insieme di espressioni semplici che trovano immediati equivalenti in altre lingue e, quindi, consentono di ricostruire le strutture concettuali di partenza in modo che siano comprensibili. Per questo motivo, come abbiamo accennato, le definizioni per parafrasi del NSM sono molto lunghe e hanno una struttura *enciclopedica* che, come sosteneva Umberto Eco (1975), dà ragione delle "differenze conoscitive".

Le ricerche di NSM hanno portato a individuare 64 *semantic primes* suddivisi in 17 categorie (*substantives; relational substantives; determiners; quantifiers; evaluators; descriptors; mental predicates; speech; actions, events, movement; location, existence, specification; possession; life and death; time; place; logical concepts; intensifier, augmentor e similarity*) e nel sito del progetto sono disponibili le tavole con le realizzazioni lessicali nelle singole lingue studiate. Per l'italiano è riportata la tavola elaborata da G.M. Farese nel 2017 (Farese 2017/2018). Nella tavola non sono presenti NIENTE o NULLA. Tra i quantificatori troviamo UNO/A~UN, DUE, TANTO, POCO~UN PO', ALCUNI, TUTTO (quindi non ZERO). Tra i *mental predicates* abbiamo VOLERE e NON VOLERE, mentre in *location, existence, specification* si trovano ESSERE (DA QUALCHE PARTE), C'E'/CI SONO,

ESSERE (QUALCUNO QUALCOSA). Tra i *logical concepts* compare la negazione NON, insieme a FORSE, POTERE, PER~PERCHE', SE.

Dunque, nel NSM *niente, nulla e zero* sono *molecole*, non elementi primitivi, ma specifici delle singole lingue. Come abbiamo accennato, il NSM ammette che esista un certo numero di molecole che siano *universal or near-universal*; si tratta di concettualizzazioni complesse, ma legate ad ambiti generali dell'esperienza umana come il corpo (*mouth, hands etc.*), l'ambiente (*sky, ground, fire, water etc.*) e i rapporti sociali (*children, woman, men etc.*).

Nell'elenco pubblicato nel sito del progetto mancano ancora *nothing o nothingness* e, quindi, la conclusione che possiamo trarre è simile a quella per parole come *God*: “the concept of ‘God’ is not universal” (Wierzbicka 2018: 21). Da oltre un ventennio⁹ Anna Wierzbicka si dedica alla definizione di *God* e delle diversità culturali legate a quella che, nella visione occidentale, è la sfera religiosa, in particolare cristiana. La definizione di *God* in termini di NSM è molto lunga, ma vogliamo riportare una parte relativa alla creazione, che è particolarmente pertinente per il nostro lavoro (Wierzbicka 2019a: 3):

The world exists because God wants it to exist

It is like this now:

There are people on earth, there are many things on earth,
many things are happening on earth.

There is the sky above the earth, there is light, there is the sun,
there is the moon, there are stars.

It is like this because God wants it to be like this.

It wasn't always like this. A very very long time ago, it was like this:

There was no earth, there was no sky, there was no light;

there were no places as there are now,

there were no people, there was nothing anywhere,

nothing was happening anywhere.

There was God.

La definizione parte dalla percezione degli oggetti esistenti e prosegue con la negazione dell'esistenza di questi oggetti nel passato; le *molecules* lessicali *nothing* e *anywhere* congiungono i *semantic primes* BE (SOMEWHERE) e BE

⁹ Cfr. Wierzbicka (2001, 2019a,b).

(SOMEONE/SOMETHING) a NOT e la definizione ri-costruisce l'atto di creazione con cui si apre l'Antico Testamento (*In principio Dio creò il cielo e la terra*) con un percorso a ritroso.

9. Veniamo alle conclusioni del nostro breve *excursus* su *niente*. Come spesso accade nella ricerca scientifica, la rassegna che abbiamo presentato ci consegna una pluralità di spiegazioni possibili. Nella visione strutturalista, il lessico dell'inesistente si dispone lungo un *continuum*, insieme semantico e sintattico. Su questa dimensione cognitiva, le parole e le combinazioni che le legano segnalano un progressivo passaggio dall'*existence* alla *nullitude*. In questa prospettiva, si conferma il valore della teoria del segno che fa dell'opposizione un principio fondante dei sistemi linguistici.

Nella costruzione ontologica di WordNet riconosciamo, da un lato, la tradizione filosofica della ricerca di categorie universali e, contemporaneamente, la prospettiva della moderna terminologia. In filosofia Bradley & Bailey (2017) ricordano Russell e Strawson che, tra gli altri, hanno postulato "a primitive of sorts", cioè una categoria che raccoglie gli elementi di tutte le altre categorie. I termini inglesi scelti per indicare questa macro-categoria sono variamente *entity*, *object*, *thing* o altri ancora, ma gli AA. concordano sulla validità ontologica di questa idea. Idea che, del resto, è presente dalla lessicografia metodica (Marello 1980) alla terminologia. Com'è noto, la Norma ISO 1087 (<https://www.iso.org/obp/ui/en/#iso:std:iso:1087:ed-2:v1:en>) parte dalla definizione di *object* come "anything perceivable or conceivable" e propone la tripartizione tra oggetti materiali (*engine*, *sheet of paper*, *diamond*), immateriali (*conversion ratio*, *project plan*) o immaginati (*unicorn*, *scientific hypothesis*).

Infine, il NSM esclude le parole dell'inesistente dall'*alfabeto del pensiero umano*. Non esistono *semantic primes* o *semantic molecules* di carattere (quasi) universale che ci presentino l'inesistente. *Niente* è, allora, come *Dio*, un concetto complesso, costruito in ambiti culturali diversi che possono essere messi in comunicazione con il ricorso a parole effettivamente universali.

Conclusioni

A nostro avviso, l'analisi linguistica offre due prospettive di lavoro interessanti. Prima di tutto, mette in evidenza l'importanza di considerare la nume-

rosità degli elementi lessicali legati all'inesistente e la varietà del loro uso. Ad esempio, *niente* è presente nei nostri discorsi come un importante indicatore referenziale (*Niente da dichiarare?*) o come un segnale che apre o mantiene la comunicazione (*Niente, ho visto Mario stamattina*). *Annientare* e *annullare* si basano su lessemi quasi-sinonimi, ma producono concettualizzazioni diverse e spesso inconciliabili (*annullare un gol* è tutt'altra cosa di *annientare un gol*, espressione che non appartiene alla comunicazione sportiva). Gli esempi in questo senso si possono moltiplicare e citiamo come esempio lo studio di De Luca (2012) dedicato alle scienze sociali in cui, partendo dalle teorie del filosofo Hans Kelsen, si discute del lessico della *nullità* che fa riferimento alle entità sociali:

Noi parliamo quotidianamente di entità sociali *nulle* (oggetti, atti, eventi, processi, stati di cose): un voto *nullo*, un contratto *nullo*, una legge *nulla* [...]. Ma non parliamo di temporali *nulli*, di schiarite *nulle* [...]. Non parliamo di una nascita *nulla* o di una morte *nulla*. (De Luca 2012: 217).

Questa varietà è spesso ritenuta un ostacolo filosofico; come afferma Reicher (2022) a proposito dell'uso di *to be* VS *to exist* "ordinary language use is too wavering and non-uniform in this respect to be a stable ground for a philosophical theory". In realtà, l'analisi della varietà degli usi linguistici apre la possibilità di un'ontologia del linguaggio naturale, che, come afferma Moltmann (2022) parte proprio dal presupposto che il linguaggio naturale rifletta l'ontologia:

the semantics of natural languages involves entities of various ontological categories, ontological structures, and ontological notions on the basis of syntactic roles of expressions, syntactic categories and features, and lexical words.

In più, l'analisi linguistica pone l'accento sulla diversità degli universi concettuali espressi dalle singole lingue. Ancora Moltmann (2022) afferma che "the ontology of natural language is the ontology speakers *implicitly* accept by way of using the language". In ambito strutturalista come in ambito cognitivo, tipicamente il lavoro descrittivo e interpretativo rivela gli elementi comuni e quelli diversi tra le lingue e rivela anche che il ragionamento e la categorizzazione sono operazioni legate alla competenza linguistica. Proprio

per questo vogliamo concludere con la citazione di un economista, Serge Wallery (2011: 86-87) che, parlando in particolare della scrittura scientifica, così si esprime:

Pendant, en sciences humaines et sociales, où la langue écrite demeure le principal véhicule de la pensée, rendre compte rigoureusement de phénomènes complexes exige une maîtrise très précise de la langue utilisée et le passage d'une langue à l'autre soulève fréquemment des problèmes de fond qui donnent lieu à d'innombrables débats de traduction – qu'on songe aux interminables querelles sur la traduction des oeuvres d'auteurs tels que Karl Marx ou Sigmund Freud.

Sia nell'argomentazione scientifica che nella comunicazione quotidiana il *contenuto* che vogliamo trasmettere “reçoit forme de la langue et dans la langue, qui est le moule de toute expression possible” (Benveniste 1958/1966: 64) e la traduzione si rivela sempre un fondamentale strumento euristico.

Riferimenti bibliografici

- Azzouni, Jodi, 2010, *Talking about Nothing: Numbers, Hallucinations, and Fictions*, Oxford, Oxford University Press.
- Benveniste, Émile, 1958, “Catégorie de pensée et catégorie de langue”, *Les Études philosophiques*, 4 (rist. in *Problèmes de linguistique générale*, Paris, Gallimard, 1966).
- Bierwisch, Manfred & Kiefer, Ferencz, 1969, “Remarks on definitions in natural language”, in Kiefer F. (ed.), *Studies in Syntax and Semantics*, Dordrecht, Reidel, 55-79
- Bradley, Rettler & Bailey, M., 2017, “Object” in Zalta E.N. (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), URL = <<https://plato.stanford.edu/archives/win2021/entries/object>>.
- Cimatti, Felice, 2013, “La zecca e l'uomo, Antropologia e linguaggio fra Wittgenstein e Lacan”, *Rivista Italiana di Filosofia del Linguaggio* 7 (2), 38-52.
- Damourette, Jacques & Pichon, Edouard, 1928, “Sur la signification psychologique de la negation en français”, *Journal de Psychologie Normale et Pathologique* 25, 228-254.
- De Mauro, Tullio, 2016, “Il Nuovo vocabolario di base della lingua italiana”, in <https://www.internazionale.it/opinione/tullio-de-mauro/2016/12/23/il-nuovo-vocabolario-di-base-della-lingua-italiana>.
- Di Lucia, Paola, 2012, “Il Nullo e il Nulla: alle radici dell'ontologia sociale”, *Rivista di Estetica*, 49, 217-225.

- Dik, Simon, 1978, *Stepwise Lexical Decomposition*, Lisse, The Peter de Ridder Press.
- Dryer, Matthews, 2013, "Indefinite Articles", in Dryer M. & Haspelmath M. (eds.) *WALS Online (2020.3) Data set [Zenodo]*, <https://doi.org/10.5281/zenodo.7385533>
- Dryer, Matthews & Haspelmath, Martin, (eds.), 2013, *WALS Online (2020.3) Data set [Zenodo]*, <https://doi.org/10.5281/zenodo.7385533>
- Eco, Umberto, 1975, *Trattato di semiotica generale*, Milano, Bompiani.
- Everett, Antony, 2013, *The Nonexistent*, Oxford, Oxford University Press.
- Fanciullacci, Riccardo, 2006, "Nulla", in *Enciclopedia filosofica Bompiani*, Milano, Fondazione Centro Studi Filosofici di Gallarate – Bompiani, 7967-7979.
- Farese, Gian Marco, 2017-2018, "The Fundamental Principles of the Italian Constitution: A Semantic Analysis", *Quaderni di Semantica* 3/4, 667-746.
- Fellbaum, Christine, 1998, (ed.), *WordNet: An Electronic Lexical Database*, Cambridge, MA, The MIT Press.
- Givone, Sergio, 2003, *Storia del nulla*, Roma-Bari, Laterza.
- Haspelmath, Martin, 2013, "Indefinite Pronouns", in Dryer M. & Haspelmath M. (eds.) *WALS Online (2020.3) Data set [Zenodo]*, <https://doi.org/10.5281/zenodo.7385533>
- Kant, Immanuel, 1959, *Critica della ragion pura*, trad. it. di Gentile G. & Lombardo Radice G. riveduta da Mathieu V., Roma-Bari, Laterza (ed. or. 1787).
- Laurent, Jérôme, 2006a, (dir.), *Cahiers de philosophie de l'Université de Caen*. 43. «Dire le néant».
- Laurent, Jérôme, 2006b, "Pourquoi parler de ce qui n'est pas?", *Cahiers de philosophie de l'Université de Caen*. 43. «Dire le néant», 7-14.
- Laurent, Jérôme & Romano, Claude, 2011a (à cure de), *Le Néant. Contribution à l'histoire du non-être dans la philosophie occidentale*, Paris, Presses Universitaires de France.
- Laurent, Jérôme & Romano, Claude, 2011b, "Préface", in Laurent, J. & Romano, C., (à cure de), *Le Néant. Contribution à l'histoire du non-être dans la philosophie occidentale*, Paris, Presses Universitaires de France, 11-32.
- Marello, Carla, 1980, *Lessico ed educazione popolare*, Roma, Armando.
- Martin, Robert, 1966, *Le mot "rien" et ses concurrents en français (du XVI^e siècle à l'époque contemporaine)*, Paris, Klincksiek.
- Martin, Robert, 1990, "La définition "naturelle", in Chaurand J. & Mazière F. (à cure de), *La définition*, Paris, Larousse, 86-95.
- Miller, George A., 1990, (ed.), *WordNet: An On-line Lexical Database, Special Issue of the International Journal of Lexicography* 3 (4).
- Miller, George A. & Johnson-Laird, Philip, 1976, *Language and Perception*, Cambridge, Cambridge University Press.

- Moltmann, Friederike, 2022, "Natural Language Ontology", in Zalta E.N. & Nodelman U. (eds.), *The Stanford Encyclopedia of Philosophy* (Winter 2022 Edition), URL = <<https://plato.stanford.edu/archives/win2022/entries/natural-language-ontology>>.
- Moro, Andrea, 2010, *Breve storia del verbo essere*, Milano, Adelphi.
- Nelson, Michael, 2022, "Existence", in Zalta E.N. & Nodelman U. (eds.), *The Stanford Encyclopedia of Philosophy* (Winter 2022 Edition), URL = <<https://plato.stanford.edu/archives/win2022/entries/existence>>.
- Nichols, S. & Stich, S., 2000, "A Cognitive Theory of Pretense", *Cognition* 74, 115-147.
- Parsons, Terence, 1980, *Nonexistent Objects*, New Haven, Yale University Press.
- Parsons, Terence, 1982, "Are There Nonexistent Objects?", *American Philosophical Quarterly* 19, 365-371.
- Peruzzi, Alberto, 1983, *Definizioni*, Milano, Angeli.
- Reicher, Maria, 2022, "Nonexistent Objects", in Zalta E.N. & Nodelman U. (eds.), *The Stanford Encyclopedia of Philosophy* (Winter 2022 Edition), URL = <<https://plato.stanford.edu/archives/win2022/entries/nonexistent-objects>>.
- Robinson, Richard, 1954, *Definition*, Oxford, Clarendon Press.
- Sager, Juan (ed.), (ed.) (2000), *Essays on Definition*, Amsterdam, Benjamins.
- Vocabolario degli Accademici della Crusca. Con tre indici delle voci, locuzioni e proverbi Latini e Greci posti per entro l'opera*, 1612, Venezia, Giovanni Alberti.
- Walery, Serge, 2011, "Productivité académique contre contribution scientifique: le cas de l'économie", *L'Économie politique* 2011 (3), 70-96.
- Wierzbicka, Anna, 1980, *Lingua mentalis*, New York, Academic Press.
- Wierzbicka, Anna, 2001, *What Did Jesus Mean? Explaining the Sermon on the Mount and the Parables in Simple and Universal Human Concepts*, Oxford, Oxford University Press.
- Wierzbicka, Anna, 2018, "Speaking about God in Universal Words, Thinking about God outside English", in Chilton P. & Kopytowska M. (eds.), *Religion, Language, and the Human Mind*, Oxford, Oxford University Press, 19-51.
- Wierzbicka, Anna, 2019b, *What Christians Believe: The Story of God and People*, Canberra, ACT, Australia, College of Arts and Social Sciences, The Australian National University, <http://doi.org/10.25911/5c4834fd34224>.
- Wierzbicka, Anna, 2019b, *What Christians Believe: The Story of God and People*, Oxford, Oxford University Press.

ELENA PEPPONI¹

Comunicazione istituzionale ampia e tecnologie per il Trattamento Automatico del Linguaggio: possibili applicazioni e sviluppi

Abstract

Al giorno d'oggi, le università sono comunità plurali, abitate da persone molto diverse tra loro: per questo motivo, esse possono essere una buona palestra di inclusione linguistica.

Questo studio, dopo aver fornito una panoramica su comunicazione istituzionale, inclusione e chiarezza, proverà ad analizzare due testi prodotti da due università italiane – e le relative riscritture inclusive – riguardanti il dominio delle carriere *alias* attraverso il tool READ-IT, valutandone leggibilità globale, profilo di base, profilo lessicale, profilo sintattico e leggibilità per frase. Lo scopo è quello di arrivare ad avanzare l'idea della creazione di un *corpus* di testi istituzionali sulle carriere *alias*. Il corpus servirà da un lato a monitorare lo stato di inclusione linguistica delle identità di genere non binarie nei testi in oggetto, e dall'altro a proporre la scrittura di testi standard, adottabili su larga scala negli atenei, in grado di coniugare le esigenze di inclusione con quelle di chiarezza e leggibilità della comunicazione istituzionale.

Parole chiave: comunicazione istituzionale, carriere *alias*, chiarezza e leggibilità, inclusione linguistica, trattamento automatico del linguaggio

Today universities are plural communities, within whom many different people live: for this reason, here our abilities for linguistic inclusion can be trained. This paper, after giving an overview about institutional communication, inclusion, and readability, will analyse, by using the tool READ-IT, two original texts by two Italian univer-

¹ Elena Pepponi, Università di Firenze, elena.pepponi@unifi.it. Per la realizzazione del presente contributo desidero ringraziare Dominique Brunato e Giulia Venturi dell'*Italian Natural Language Processing Laboratory* (ItalianNLP Lab) interno all'Istituto di Linguistica computazionale Antonio Zampolli del CNR di Pisa: loro mi hanno insegnato a conoscere e ad usare READ-IT, suggerendomi una ricca bibliografia in merito, indispensabile per la scrittura di questo articolo. Il mio ringraziamento va anche a Gloria Comandini (Istituto Italiano di Studi Germanici – Roma), che ha accettato con molto entusiasmo di dividere con me l'oneroso compito di creare e gestire il *corpus ALIAS*, di cui si parlerà nel § 4. Naturalmente, ogni imperfezione che dovesse emergere dalla stesura del contributo è imputabile soltanto a me.

sities regarding *alias* identities, and their re-written versions. With the help of this tool, global readability, basic profile, lexical profile, syntactical profile and sentence readability of two institutional texts about *alias* identities will be evaluated. The aim of the research is to suggest the creation of a *corpus* of institutional texts about *alias* identities in Italy. The *corpus* will be useful to track the current linguistic inclusion of non-binary people; at the same time, after quantitative analysis, it will be possible to write standard and shareable versions of institutional texts about *alias* identities, able to integrate strategies for gender inclusion with strategies for clear and readable institutional communication.

Keywords: institutional communication, *alias* identities, openness and readability, linguistic inclusion, natural language processing

1. Introduzione

È ormai un dato di fatto che nella società odierna ci siano persone che non si riconoscono, del tutto o in parte, nel rigido sistema di genere a due opzioni che per molto tempo è stato stabilito come normativo. Queste persone, per quanto presenti in un numero relativamente basso, sono, come tutte le altre, fruitrici di quei testi e quei prodotti che rientrano all'interno del termine ombrello 'comunicazione istituzionale', e come tali hanno il pieno diritto di vedersi proporre dei testi che siano costruiti con un linguaggio il meno discriminatorio possibile.

Le istituzioni, e gli atenei in particolare, si stanno già muovendo da diverso tempo nella direzione di una comunicazione sensibile alle differenze di genere, specialmente grazie al costante lavoro dei CUG (Comitato Unico di Garanzia per le pari opportunità, la valorizzazione del benessere di chi lavora e contro le discriminazioni); tuttavia, fino a non molto tempo fa, la prospettiva di inclusione anche linguistica era perlopiù solo binaria, cioè si preoccupava di dover includere uomini e donne, non altre casistiche. Come però fa notare Thornton (2022: 49), “[l]a necessità di dare rappresentanza a persone con identità di genere non allineata su uno dei due poli che costituiscono il nucleo semantico dei due valori di genere oggi presenti in italiano standard è [...] più difficile”.

Nel presente contributo si cercherà quindi di dare una parziale risposta alle seguenti domande: come si sta aprendo la comunicazione istituzionale

universitaria alla rappresentazione di nuove identità di genere non binarie? Si può tentare di veicolare un'informazione attraverso un linguaggio che cerchi di essere poco o per nulla discriminatorio senza allo stesso tempo snaturare la comunicazione istituzionale stessa? Come possono essere impiegati gli strumenti di Trattamento Automatico del Linguaggio per monitorare testi istituzionali e per proporre modifiche?

Prima di tutto (§ 2), si condurrà una panoramica preliminare su quante e quali siano le strategie linguistiche che possono garantire una comunicazione istituzionale davvero inclusiva, e allo stesso tempo si rifletterà su come esse possano entrare in rapporto con altre strategie, in particolare quelle di chiarezza, semplicità e fruibilità dei testi. I testi analizzati riguarderanno l'ambito delle carriere *alias*, quelle identità transitorie e non consolidabili che possono essere attribuite all'interno delle università a persone *gender variant*² al fine di consentire loro di vivere la vita accademica conformemente al proprio genere percepito ed essendo riconosciute con il proprio nome eletto. Queste nuove identità si sovrappongono alla carriera già maturata fino a quel momento - in caso di persona studente - o all'anagrafica dell'individuo in caso di personale di ricerca, docente o tecnico-amministrativo.

In secondo luogo (§ 3), attraverso l'uso di strumenti di Trattamento Automatico del Linguaggio, si analizzeranno due testi istituzionali realmente esistenti e le loro riscritture qui proposte, realizzate proprio sulla base delle riflessioni teoriche contenute nel § 2. Un'Appendice finale, organizzata in due colonne, renderà immediatamente visibili sulla colonna di sinistra i testi originali e su quella di destra le modifiche apportate in questa sede con criteri inclusivi di genere plurale. In particolare, grazie a READ-IT, un *tool* sviluppato presso l'Istituto di Linguistica Computazionale *Antonio Zampolli* del CNR di Pisa per annotare e valutare testi in italiano, di essi si analizzerà la leggibilità della versione originale e si comparerà con quella della versione riscritta dai seguenti punti di vista: leggibilità globale del testo, profilo di base, profilo lessicale, profilo sintattico e *sentence readability*.

² Per *gender variant* si intendono tutte quelle persone che non si riconoscono nel genere che viene loro attribuito alla nascita e/o che contestano l'impostazione binaria del concetto stesso di genere.

Infine (§ 4), si proporrà la creazione di un *corpus* monolingue di tutti i testi delle università italiane riguardanti le carriere *alias*, che possa fungere come bacino di monitoraggio per la comunicazione istituzionale orientata all'inclusione di genere plurale e che, sperabilmente, possa favorire la stesura di testi che funzionino come modelli testuali standard per le università su base nazionale.

2. Quali strategie per una comunicazione istituzionale ampia e chiara?

Le strategie per costruire una comunicazione istituzionale chiara e leggibile a prescindere dalla provenienza e dal grado di scolarizzazione dell'utenza sono oggetto di studi sin dagli anni Ottanta del secolo scorso³. Una sintesi di tutte le buone pratiche che si possono adottare per produrre testi semplici, chiari ed efficaci è quella rappresentata dalle *Trenta regole per scrivere testi amministrativi chiari* (Cortelazzo & Pellegrino 2002), riprese anche in Cortelazzo (2021: 144 e ss.).

Inoltre, da più di vent'anni si ragiona su come costruire una comunicazione pubblica che sia anche rispettosa delle differenze di genere, cioè che sia in grado di fornire una corretta rappresentazione sia del genere maschile sia di quello femminile, ove coinvolti nell'interazione. Numerose figure di studiose e studiosi si sono dedicate, anche in questo caso, alla produzione di studi teorici e linee guida pratiche per la scrittura istituzionale sensibile alle differenze di genere⁴. Poiché negli ultimi anni stanno iniziando ad affermarsi le rivendicazioni di tutte quelle persone che non si riconoscono, del tutto o in parte, nella casistica 'uomo' o 'donna', può essere interessante iniziare a

³ Per una panoramica su chiarezza e semplificazione della comunicazione istituzionale cfr. almeno Bombi (2013; 2015), Cassese (1983), Cortelazzo (1999; 2014a; 2014b; 2015; 2021), Cortelazzo & Pellegrino (2003), Fioritto (1997), Franceschini & Gigli (2003), Lubello (2014; 2017), Vellutino (2018), Viale (2008).

⁴ Oltre ai numerosi lavori della principale studiosa del settore, Cecilia Robustelli (cfr. almeno Robustelli 2012; 2015; 2016; 2021), di rilievo è anche la presa di coscienza di diversi atenei italiani sull'importanza della visibilità linguistica del genere, sfociata poi in numerose linee guida e diversi consigli pratici per la realizzazione di una scrittura istituzionale rispettosa delle differenze. A questo proposito, si vedano ad esempio le *Linee guida per la visibilità del genere nella comunicazione istituzionale dell'Università di Bologna*, pubblicate nel 2020 (<[https://www.unibo.it/it/ateneo/chi-siamo/linee-guida-per-la-visibilita-del-genere-nella-comunicazione-istituzionale-universita-di-bologna](https://www.unibo.it/it/ateneo/chi-siamo/linee-guida-per-la-visibilita-del-genere-nella-comunicazione-istituzionale)>, ultima consultazione: settembre 2023), e l'opuscolo informativo *Generi e linguaggi. Linee guida per un linguaggio amministrativo attento alle differenze di genere*, promosso dall'Università di Padova nel 2017 (<<https://www.unipd.it/sites/unipd.it/files/2017/Generi%20e%20linguaggi.pdf>>, ultima consultazione: settembre 2023).

esplorare nuove soluzioni per la comunicazione istituzionale che siano rappresentative delle persone a prescindere dalla loro identità di genere, ma che allo stesso tempo cerchino di integrare gli sforzi fatti finora per la visibilità binaria e anche quelli per la comunicazione chiara e semplice.

Da un punto di vista binario, le tecniche per la corretta espressione linguistica del genere nei testi sono riconducibili a due macro-tipologie: quelle orientate alla modifica dei tratti morfosintattici e quelle incentrate sui tratti lessico-frasali.

Fanno parte delle prime quegli accorgimenti grammaticali che mirano ad abbattere l'uso del cosiddetto *maschile sovraesteso*, producendo invece accordo tra le varie parti del discorso in modo corretto secondo il genere della referenza extralinguistica. Se queste tecniche hanno funzionato egregiamente per la comunicazione inclusiva in prospettiva binaria, esse risultano però limitate nel momento in cui si vogliono rappresentare correttamente altre identità di genere. In italiano il genere grammaticale prevede infatti soltanto la possibilità di un maschile e di un femminile: le varie proposte di sperimentazione per l'adozione di una terza casistica, se pur molto fertili in diversi contesti⁵, sembrano ancora molto lontane dal raggiungere, a oggi, un apprezzabile livello di normatività nell'italiano standard che possa garantire il loro impiego nella comunicazione istituzionale. Sebbene alcuni studiosi e alcune studiose stiano iniziando a occuparsene, infatti, non abbiamo al momento degli studi completi che ci spieghino se e come queste strategie possano essere processate dall'utenza finale, e se eventualmente possano rappresentare un maggior aggravio cognitivo nella fruizione testuale⁶.

Sono invece strategie lessico-frasali quelle orientate alla neutralizzazione di genere da un punto di vista non strettamente morfologico: tra di esse si an-

⁵ A questo proposito rinvio al lavoro di Comandini (2021), che indaga l'uso di varie strategie morfosintattiche di neutralizzazione di genere (ə, *, u, x e @).

⁶ Per il tedesco è stato condotto nel 2021 uno studio atto a verificare se l'uso dell'asterisco come morfema per indicare persone la cui identità di genere non è nota abbassasse o meno il livello di riconoscibilità delle parole (cfr. Kruppa, Fenn & Ferstl 2021): i risultati di tale studio, portato a termine attraverso una *lexical decision task* sottoposto a un campione di persone, sembrano suggerire che la capacità di riconoscimento di parola non venga intaccata in maniera significativa dall'uso dell'asterisco. Per l'italiano non abbiamo, a oggi, studi paragonabili che indaghino la capacità dei e delle parlanti di comprendere un testo in cui la neutralizzazione di genere sia affidata a simboli come <*> o <ə>: pertanto, non siamo in grado di affermare che essi siano certamente una fonte di complicazione testuale, ma nemmeno di sostenere il contrario.

noverano l'uso di epiceni e di nomi di genere comune⁷, l'uso di forme impersonali, la costruzione passiva, l'uso del pronome indefinito *chi* etc. Alcune di queste strategie sembrano confliggere con le *Trenta regole* prima menzionate in termini di creazione di testi chiari, come illustrato nella tabella 1.

Strategia lessico-frasale di neutralizzazione di genere	<i>Trenta regole per scrivere testi amministrativi chiari</i>
Uso di frasi passiva	Regola 18 - <i>Preferite la forma attiva</i>
Uso di forme impersonali, uso del pronome indefinito <i>chi</i>	Regola 19 - <i>Evitate le frasi impersonali</i>
	Regola 20 - <i>Evitate la nominalizzazione</i>
	Regola 21 - <i>Esprimate il soggetto</i>

Tabella 1. *Strategie lessico-frasali per la neutralizzazione di genere a confronto con le Trenta regole di Cortelazzo & Pellegrino (2002).*

Nonostante ciò, oggi iniziano ad essere comunque in atto diversi tentativi per realizzare una comunicazione allo stesso tempo inclusiva e leggibile, tentativi che però stanno procedendo ‘a macchia di leopardo’ a seconda della maggiore o minore sensibilità sul tema espressa dalle diverse istituzioni nazionali o locali, e che non possono ancora dirsi del tutto incisivi o condivisi. Tuttavia, se si vuole uscire dall'*impasse* binaria, l'unica strada che sembrerebbe percorribile sarebbe quella di implementare le strategie finora utilizzate per allargare le maglie della rappresentatività, e le uniche flessibili e passibili di miglioramento sono, a oggi, le strategie lessico-frasali. Una possibile implementazione che non entri in conflitto con le *Trenta regole* è l'uso di epiceni come *persona* e *individuo*, accompagnati, quando necessario, da una componente attributiva che li completi. Scrivendo, ad esempio, *persona studente*, *persona laureanda* o *persona docente* rispetto a *studente/studentessa*, *laureando/laureanda*, *il/la docente* o *il professore/la professoressa*, si evita di ricorrere a un'impostazione solo binaria del linguaggio. Allo stesso tempo, preferendo la formula con *persona* a eventuali formule come *chi studia*, *chi si laurea/*

⁷ Uso in questa sede i termini *epiceni* e *nomi di genere comune* nell'accezione proposta da Thornton (2022: 21). Gli *epiceni* sono dunque nomi che “hanno un'unica forma e presentano target di accordo di un unico genere, ma possono designare individui di qualunque sesso (per es. *la persona*)”, laddove i *nomi di genere comune* sono quelli che “hanno un'unica forma ma presentano target di accordo dei due generi, a seconda del sesso del referente (per es. *il cantante* VS. *la cantante*)”.

deve laurearsi, chi insegna, si può evitare l'uso della frase passiva e del pronome indefinito, noti elementi di complicazione testuale. Questa sembrerebbe dunque essere una buona strategia per coniugare l'esigenza - ormai sempre più attuale - di includere dal punto di vista linguistico le persone che non si riconoscono nel binarismo di genere con quella di produrre testi che non siano in aperta contraddizione con l'ormai consolidata letteratura sulla comunicazione pubblica chiara e semplice, né con le linee guida proposte nel corso del tempo per realizzarla concretamente. Va però sottolineato come questa soluzione richiederebbe un compromesso da parte di quelle rivendicazioni per la visibilità del femminile, faticosamente raggiunta nella comunicazione istituzionale - e tutto sommato ancora non totalmente pacifica nonostante il tempo trascorso -, e che ora rischia di trovarsi nuovamente opacizzata da una neutralizzazione onnicomprensiva⁸.

Nei prossimi paragrafi vedremo come applicare questi possibili accorgimenti a testi reali, facendo anche uso di tecnologie di Trattamento Automatico del Linguaggio per la fase di monitoraggio dei risultati ottenuti.

3. Testi istituzionali e strumenti di Trattamento Automatico del Linguaggio: il caso di studio

In questo paragrafo, a partire dalle riflessioni condotte in quello precedente, si faranno degli esperimenti pratici, ponendosi le seguenti domande:

- la riscrittura di testi istituzionali con accorgimenti linguistici rappresentativi di identità non solo binarie può inserirsi in una testualità che mostra già, come sappiamo, dei suoi profili di complessità?

⁸ Di questo fragile equilibrio tra istanze del femminile e istanze non binarie parla diffusamente Thornton (2022: 46 ss.) nel paragrafo conclusivo del suo lavoro, intitolato *È possibile un mutamento linguistico volontario?*. Naturalmente, il dominio di testi qui analizzati, ovvero quelli per le carriere *alias*, è particolare, in quanto si tratta di testi dedicati a esigenze specifiche di persone *gender variant*: è dunque più facile che il compromesso - per così dire - 'a ribasso' per le esigenze di visibilità linguistica del femminile sia accettato senza grandi opposizioni, dato che si dà, o si dovrebbe dare, priorità alle persone di genere non binario, *transgender* o *queer*. Il discorso potrebbe non essere lo stesso nel caso di testi di istituzioni maggiori rivolti a una moltitudine di persone più vasta ed eterogenea: in quel caso, per statistica, le persone *gender variant* sarebbero quasi sicuramente una ridotta minoranza, dunque potrebbero esserci più resistenze da parte della popolazione *cisgender* femminile a cedere porzioni di visibilità in nome di un'inclusione a-specifica.

- tali accorgimenti, com'è spesso timore di chi deve scrivere o riscrivere, complicano ulteriormente i testi abbassandone il grado di leggibilità?

Per dare una risposta a tali domande, sono stati selezionati come campione due testi diversi tra loro sia per tipologia che per canale di fruizione, entrambi riguardanti il dominio delle carriere *alias*: a) la pagina del sito *web* dell'Università di Firenze dedicata alle carriere *alias* ove, attraverso un formato Q&A, si spiega come richiedere tale servizio; b) il Regolamento per l'attivazione delle carriere *alias* approvato dall'Università di Pisa nel 2020⁹. Lungo l'articolo, per ragioni di chiarezza e brevità, ci si riferirà ai testi in oggetto con le seguenti sigle:

- testo originale del sito *web* dell'Università di Firenze: CAUnifi-Or;
- testo riscritto del sito *web* dell'Università di Firenze: CAUnifi-Ri;
- testo originale del Regolamento dell'Università di Pisa: CAUnipi-Or;
- testo riscritto del Regolamento dell'Università di Pisa: CAUnipi-Ri;

Secondo la classificazione proposta da Viale (2008: 109)¹⁰, CAUnifi-Or e CAUnifi-Ri appartengono alla tipologia testuale 'testo per pagine web', pertinente alla funzione descrittiva; dal punto di vista delle persone destinatarie, essi possono essere considerati testi rivolti all'esterno ma con forte rilevanza interna. Diversamente, i testi CAUnipi-Or e CAUnipi-Ri appartengono alla tipologia 'regolamento', hanno una vocazione spiccatamente prescrittiva e si trovano all'interfaccia tra testi rivolti all'interno con forte rilevanza esterna e testi rivolti all'esterno con forte rilevanza interna.

Di entrambi i testi è stata fatta una prova di riscrittura utilizzando le strategie lessico-frasali di rappresentatività linguistica implementate secondo le riflessioni contenute nel § 2, e riformulando vari passaggi ove non chiari o

⁹ Cfr. rispettivamente <<https://unifinclude.unifi.it/vp-93-la-carriera-alias.html>> e <<https://www.unipi.it/phocadownload/regolamentiallegati/regcaralias.pdf>> (ultima consultazione di entrambi: settembre 2023).

¹⁰ In questo luogo l'autore riprende e integra la classificazione già avanzata da Mortara Garavelli (1997 [1988]).

scorretti¹¹. Non sono, però, state applicate sistematicamente le *Trenta regole* né altre strategie mirate esplicitamente alla semplificazione e alla chiarezza. In questo contesto, infatti, si vuole solo attirare l'attenzione sulla possibilità di scrivere in modo rappresentativo di diverse identità di genere senza esplicitamente dedicarsi a migliorare il livello di chiarezza dei testi, basso o alto che sia, che le università hanno deciso di esprimere negli originali. Sulla possibilità di integrare le strategie lessico-frasali di rappresentazione di genere con le istanze della comunicazione istituzionale chiara, semplice ed efficace si discuterà poi nel § 4.

Successivamente, sia i testi originali che i testi riscritti sono stati analizzati con l'ausilio di READ-IT, uno strumento di monitoraggio della leggibilità dei testi italiani¹². A partire dai parametri impostati su READ-IT si verificherà la leggibilità dei testi di partenza dal punto di vista del profilo di base, di quello lessicale e di quello sintattico, si cercherà di capire se i testi in questione sono stati redatti in modo leggibile e se, e in quale misura, gli accorgimenti linguistici inclusivi possano modificare questo tasso di leggibilità.

3.1 *Panoramica generale e profilo di base*

READ-IT è uno strumento “utilizzato per valutare l'efficacia comunicativa di testi in diverse tipologie di comunicazione” (Brunato & Venturi 2014: 116) che però, rispetto ad altri metodi di misurazione della leggibilità come l'indice Gulpease¹³, non si basa sulle sole caratteristiche formali e generali del testo - come ad esempio la lunghezza della frase e quella della parola - ma “implementa un indice di leggibilità ‘avanzato’ basato su analisi linguistica multi-livello del testo condotta con strumenti che rappresentano lo stato dell'arte per il trattamento automatico della lingua italiana” (Brunato &

¹¹ Si precisa infatti che le riscritture possono interessare la forma, ma talvolta anche la sostanza, ove incorrono inesattezze nel trattamento linguistico delle persone non binarie, sulla base delle linee guida fornite da diversa documentazione giuridica e sanitaria. Si possono visionare i testi e le relative riscritture nell'Appendice.

¹² Cfr. <<http://www.italianlp.it/demo/read-it/>> (ultima consultazione: settembre 2023). Per una panoramica sullo strumento READ-IT, sui suoi parametri d'impostazione e sul suo funzionamento cfr. Dell'Orletta, Montemagni & Venturi (2011).

¹³ Per un approfondimento sull'indice Gulpease cfr. Lucisano & Piemontese 1988.

Venturi 2016: 131)¹⁴. READ-IT, dunque, oltre a permettere la misurazione dei parametri formali, consente di annotare il testo in una prospettiva lessicale, morfo-sintattica e sintattica, nonché di visualizzare le proiezioni della leggibilità sulle singole frasi che lo compongono.

In questo paragrafo, come suggerisce il suo titolo, ci focalizzeremo sulle caratteristiche generali (READ-IT Base, READ-IT Globale, Gulpease) e sul profilo di base dei testi tanto originali quanto riscritti, demandando ai successivi paragrafi le riflessioni lessicali e sintattiche.

Nelle figg. 1 e 2 possiamo osservare la panoramica generale e il profilo di base rispettivamente di CAUnifi-Or e di CAUnifi-Ri.

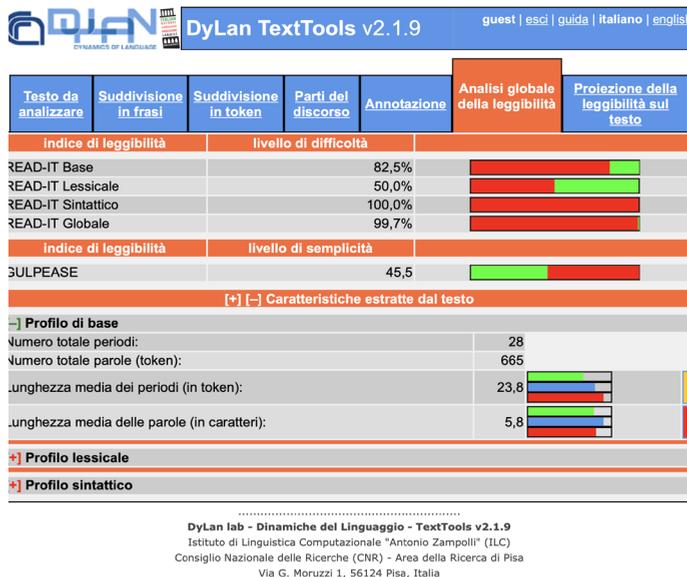


Figura 1. Panoramica generale e profilo di base di CAUnifi-Or.

¹⁴ Oltre ai già citati Dell'Orletta, Montemagni & Venturi (2011) e Brunato & Venturi (2014), cfr. anche Brunato (2014); Dell'Orletta, Montemagni & Venturi (2014); Montemagni (2013). È bene ricordare che READ-IT misura il livello di difficoltà di un testo, mentre l'indice Gulpease è un indicatore di facilità, basato quindi su una proporzione inversa: più alto è il punteggio, più basso è il livello di scolarizzazione necessario per avvicinarsi al testo in autonomia. Nonostante la sua maggiore specificità, tuttavia, READ-IT include anche il risultato fornito dal Gulpease.

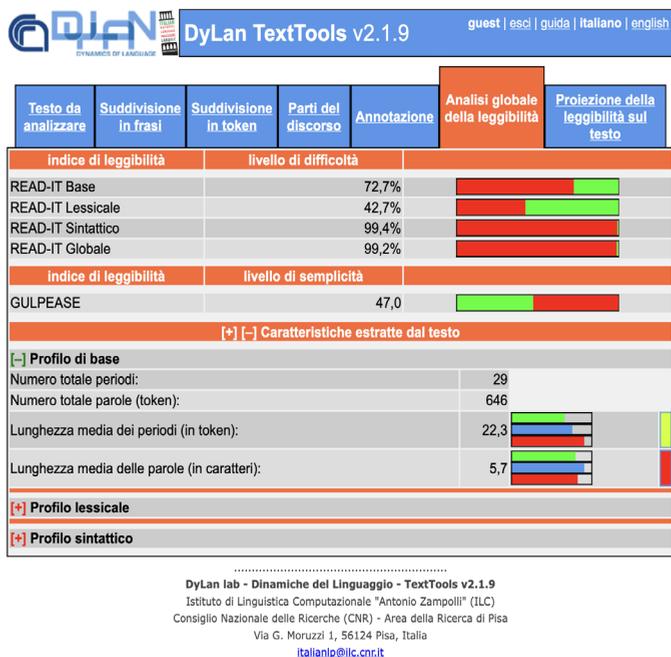


Figura 2. Panoramica generale e profilo di base di CAUnifi-Ri.

CAUnifi-Or è molto complesso: l'indice READ-IT Base pone un livello di difficoltà dell'82,5%, l'indice READ-IT Lessicale del 50% e quello Sintattico del 100%. Il READ-IT Globale, che fornisce per l'appunto una visione generale combinata dei tassi di complessità di base, lessicali, morfo-sintattici e sintattici, attesta la difficoltà del testo sul 99,7%. È inequivocabilmente un testo complesso da leggere e comprendere, come corroborato anche dall'indice Gulpease, che segna 45,5 punti: secondo questo parametro, il prodotto testuale è leggibile da persone che abbiano un livello di scolarizzazione non inferiore al diploma di scuola secondaria superiore.

Tuttavia, contrariamente a quanto il pregiudizio sulla comunicazione inclusiva fa temere, i dati visibili in fig. 2 ci suggeriscono che, per alcuni parametri, le strategie di riscrittura ampia e non binaria sono tutt'altro che

responsabili di un'ulteriore complicazione del testo. Infatti, in CAUnifi-Ri il livello di difficoltà del READ-IT Base scende di quasi dieci punti percentuali (72,7%) e quello Lessicale di più di sette (42,7%), mentre quello Sintattico ha una lievissima oscillazione non significativa verso il basso, attestandosi sul valore di 99,4%. Non significativa è anche l'oscillazione del READ-IT Globale nella riscrittura (99,2%), ma ciò non deve stupire: la panoramica globale, infatti, tiene conto di tutti i parametri, e vedremo nei paragrafi successivi che le difficoltà soprattutto morfosintattiche e sintattiche non vengono modificate in maniera importante dagli interventi per l'inclusione di genere. Se pur poco apprezzabile, si nota in CAUnifi-Ri anche un piccolo miglioramento dell'indice Gulpease (47,0 punti), che tuttavia rimane ampiamente sotto la soglia dei 60 punti, quindi ancora scarsamente o per nulla leggibile in autonomia da persone che hanno una scolarizzazione inferiore alla quinta superiore.

Approfondendo il profilo di base, notiamo che il numero totale di periodi rimane quasi identico (28 in CAUnifi-Or, 29 in CAUnifi-Ri), mentre ha una lieve variazione in negativo il numero di *token* (665 in CAUnifi-Or, 646 in CAUnifi-Ri). La lunghezza media dei periodi in *token* diminuisce di 1,5 (23,8 in CAUnifi-Or, 22,3 in CAUnifi-Ri)¹⁵, mentre quella delle parole in caratteri ha una diminuzione minima e non significativa (5,8 in CAUnifi-Or, 5,7 in CAUnifi-Ri).

In questo caso, dai dati non si apprezzano dunque modifiche che possano scoraggiare le riscritture inclusive come possibile fonte di complicazione del testo, anzi, come visto, in alcuni casi ci sono persino dei miglioramenti di alcuni valori di leggibilità.

Andremo ora a proporre la medesima riflessione per CAUnipi-Or e per CAUnipi-Ri, i cui dati possono essere visionati nelle figg. 3 e 4.

¹⁵ A proposito della lunghezza media dei periodi, è interessante qui notare come il tool, impostato su rigide regole numeriche, faccia apparire una luce verde chiaro per CAUnifi-Ri, mentre la luce è arancione, quindi indice di un testo più complesso, per lo stesso parametro in CAUnifi-Or.

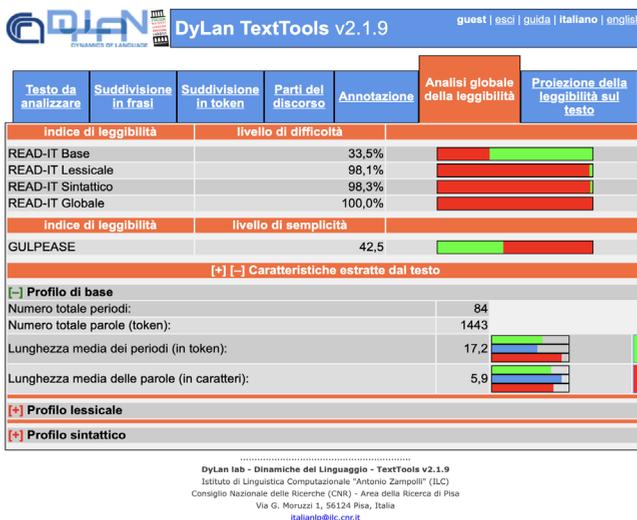


Figura 3. Panoramica generale e profilo di base di CAUnipi-Or.

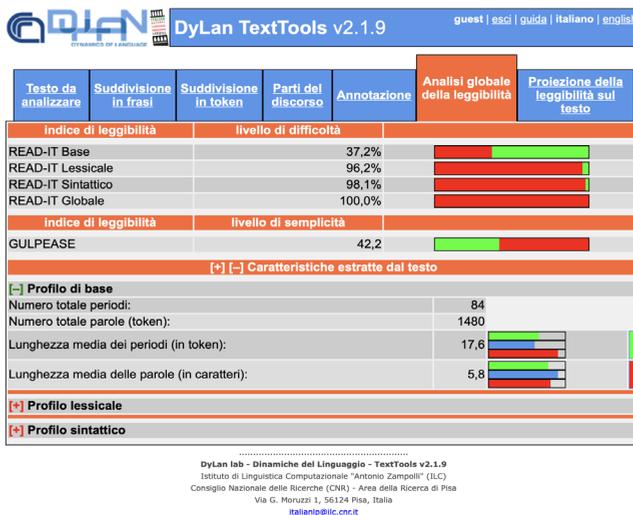


Figura 4. Panoramica generale e profilo di base di CAUnipi-Ri.

Il READ-IT Base di CAUnipi-Or ha un valore molto basso (33,5%), mentre sono elevatissimi i valori del READ-IT Lessicale (98,1%) e di quello Sintattico (98,3%). Questi parametri fanno arrivare il valore READ-IT Globale al 100% della difficoltà; l'indice Gulpease è a 42,5 punti, perciò anche in questo caso il testo è leggibile in autonomia da persone con almeno la licenza di scuola superiore.

In CAUnipi-Ri, come vediamo, aumenta il livello di difficoltà del READ-IT Base di quasi 4 punti percentuali (37,2%) rispetto a CAUnipi-Or, mentre l'indice di leggibilità Lessicale diminuisce di quasi due punti rispetto all'originale (96,2%) e quello Sintattico ha una diminuzione non significativa (98,1%). Si mantiene invece inalterato sul 100% quello del READ-IT Globale; non significativa, anche se presente, la variazione dell'indice Gulpease (42,2 punti).

Ci si potrebbe domandare come mai in un testo a evidente vocazione tecnica come CaUnipi-Or il READ-IT Base presenti un valore molto basso, tipico di testi descrittivi e divulgativi, e perché, invece, paradossalmente, la sua riscrittura ne presenti uno più elevato.

Per provare a rispondere alla domanda bisogna ricordare che, come detto in precedenza, il READ-IT Base tiene in considerazione parametri estremamente basilari quali lunghezza delle parole e lunghezza dei periodi. Tuttavia, spiega Brunato (2014: 18), quello della lunghezza dei periodi è “un parametro molto grezzo rispetto alla caratterizzazione della complessità di un testo, nonostante sia proprio quello più utilizzato per formalizzare il grado di complessità sintattica degli indici di leggibilità tradizionali”. Se è vero che “la lunghezza dei periodi costituisce forse la caratteristica dello stile amministrativo che balza più immediatamente agli occhi, e quanto più una frase è lunga, maggiore è la probabilità che sia sintatticamente complessa” (Fortis 2005: 65), è altrettanto vero che, per l'appunto, una tendenza non è una certezza, e che possono rilevarsi luoghi di complessità in testi con parole corte e periodi corti tanto quanto luoghi di semplicità in testi con periodi e *token* molto lunghi. Non è strano, quindi, che un testo come CaUnipi-Or presenti un valore molto basso di READ-IT Base che sembra andare in contrasto con i valori Sintattico e Lessicale, entrambi elevatissimi: ci accorgeremo infatti più avanti della profonda complessità del testo nonostante il valore basso di READ-IT Base.

Per quanto riguarda l'aumento del valore READ-IT Base nella riscrittura rispetto all'originale, vediamo che in CAUnipi-Or il numero totale di periodi è 84 e il numero totale dei *token* è 1443. La lunghezza media dei periodi in *token* è di 17,2, mentre la lunghezza media delle parole in caratteri è di 5,9.

Per quanto riguarda CAUnipi-Ri, si nota che, a fronte dello stesso numero di periodi (84), il numero totale di *token* è aumentato di 37 unità (1480), ma la lunghezza media dei periodi in *token* è rimasta molto simile (17,6), come pure la lunghezza media delle parole in caratteri (5,8).

Un'ipotesi di risposta potrebbe trovarsi nella tecnica di *splitting* di genere utilizzata in CAUnipi-Or. Questo testo, infatti, usa la doppia forma maschile/femminile, ma le forme femminili sono spesso apocotate e non complete (per es. *del/la, studente/ssa*). Esse sono state modificate in CAUnipi-Ri con alternative inclusive (per es. *persona studente*), tranne nel caso di *il/la o del/la Rettore/trice* che è stato modificato semplicemente in *il o del Rettore e la o della Rettrice*¹⁶, eliminando dunque le apocopi. Perciò, è possibile supporre che i gruppi di lettere che nella versione originale comparivano dopo la barra diagonale non fossero considerati parole autonome, mentre invece lo diventino nella versione riscritta, quando assumono appunto uno statuto lessicale completo. Questo spiegherebbe l'aumento del numero totale di *token* del testo, e, anche se poi i valori medi non sono cambiati quasi per nulla, ciò potrebbe aver spinto lo strumento automatico a considerare il testo un po' più complesso, anche se questa informazione non necessariamente corrisponde alla verità. Vedremo grazie alle analisi lessicali e sintattiche svolte nei prossimi paragrafi come il parametro della lunghezza dei *token* e delle frasi sia assolutamente insufficiente, da solo, a misurare e restituire i livelli di complessità dei testi.

3.2 Profilo lessicale

In questo paragrafo ci occuperemo di monitorare i testi originali e le riscritture dal punto di vista del loro profilo lessicale.

Tre sono le caratteristiche che il *tool* prende in considerazione all'interno dei testi analizzati. Innanzitutto, la distribuzione del lessico rispetto al Vocabolario di Base (VdB), così come formalizzato nel *GRADIT* di Tullio De Mauro, con conseguente ripartizione dei *token* rispetto ai tre repertori che costitui-

¹⁶ A oggi, in Italia, non esistono persone *gender variant* che ricoprono la carica del Rettorato, esistono solo Rettori o Rettrici: non si ravvisa, quindi, l'esigenza di un ipercorrettismo inclusivo di genere non binario. Naturalmente, qualora dovesse presentarsi in futuro questa eventualità, si potrà applicare lo stesso principio usato per *persona studente o persona docente* e riferirsi a tale carica con l'appellativo di *Persona Rettrice*.

scono il VdB stesso, ossia il lessico fondamentale (F), quello di alto uso (AU) e quello di alta disponibilità (AD). In secondo luogo, il rapporto tipo/unità (noto come *type/token ratio* o TTR). La TTR è

la misura che mette a rapporto il numero delle occorrenze di unità del vocabolario di un testo (al denominatore) con il numero di parole tipo (al numeratore): il suo valore oscilla nell'intervallo tra 0 e 1. Si tratta di un parametro che formalizza la varietà lessicale di un testo e che tende ad aumentare all'aumentare della complessità di un testo (Brunato 2014: 13)¹⁷.

Infine, la densità lessicale, ovvero il rapporto tra parole piene e totale delle occorrenze di parole: ancora secondo Brunato (2014: 14), “la letteratura sulla leggibilità dimostra che valori superiori di densità lessicale sono associati ad un maggior carico informativo e, dunque, a una maggiore complessità”.

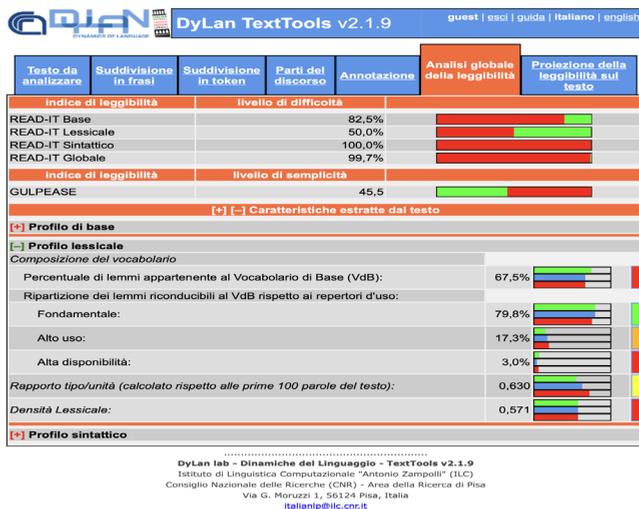


Figura 5. Profilo lessicale di CAUnifi-Or.

¹⁷ Nel caso dei testi che si stanno prendendo in considerazione in questo contributo il problema della diversa misura quasi non si pone, dato che, come abbiamo visto, le lunghezze sono più o meno uguali tra testo originale e testo riscritto, trattandosi di materiale testuale parallelo (cfr. § 4). Tuttavia, potrebbero essere analizzati testi di lunghezze molto diverse, dunque la TTR è convenzionalmente misurata rispetto ai primi 100 *token* del testo.

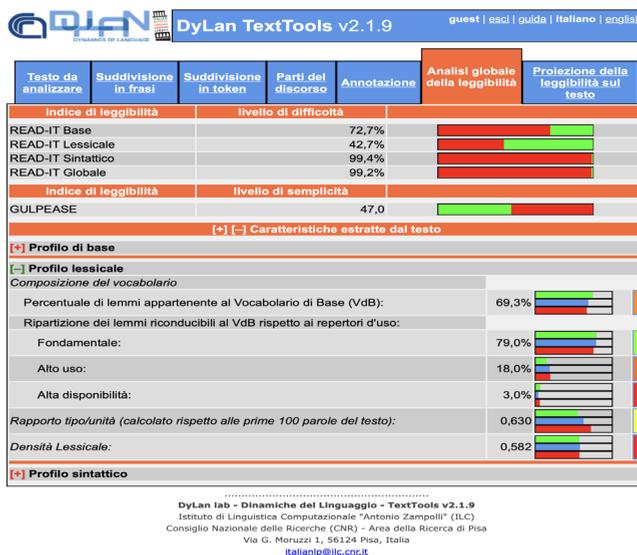


Figura 6. Profilo lessicale di CAUnifi-Ri.

Come possiamo notare, in CAUnifi-Or (fig. 5) la percentuale di lemmi che appartengono al VdB è del 67,5%, il che qualifica il testo come complesso. Se da una parte questo risultato non stupisce, trattandosi di un prodotto testuale istituzionale, tuttavia esso getta una luce piuttosto fosca sui tentativi che le università fanno di essere chiare e comprensibili: ricordiamo infatti che CAUnifi-Or è tratto da una pagina di sito *web* organizzata secondo il modello Q&A, quindi teoricamente non è pensato per essere un testo tecnico, ma è stato ideato per la divulgazione, eppure usa una percentuale non elevata di VdB. Nella ripartizione dei lemmi rispetto ai tre repertori del VdB, notiamo la grande preponderanza del lessico F rispetto a quelli di AU e AD (F = 79,8%; AU = 17,3%; AD = 3,0%). Il comportamento di CAUnifi-Or per quanto riguarda la distribuzione del VdB è quindi più vicino a quello di un testo tecnico che a quello di uno divulgativo.

È invece interessante notare i dati sulla TTR. Abbiamo detto poco sopra che la TTR aumenta all'aumentare della complessità testuale, e in effetti nel misurare questo valore bisogna sempre tenere conto del genere testuale. In

presenza di un testo più tecnico il rapporto tipo/unità si alza, e ciò è facilmente spiegabile. Fa notare infatti Brunato (2014: 13), riprendendo le riflessioni di Goti (2005: 33), che i testi più tecnici, per non parlare di quelli fortemente specialistici, hanno una scarsa varietà lessicale perché non possono fare ricorso con leggerezza alla sinonimia, dato che la terminologia è monoreferenziale, perlomeno all'interno del dominio di specialità. Ecco perché essi presentano valori di TTR più alti rispetto ai testi divulgativi. Anche se questo testo ha dei luoghi di complessità che sicuramente lo allontanano dal valore divulgativo che avrebbe voluto esprimere, non è però un testo eminentemente tecnico come può essere ad esempio un regolamento. Con il valore di 0,630 e una luce gialla - che segnala testi di difficoltà intermedia - la TTR di CAUnifi-Or è infatti elevata, ma tutto sommato meno vicina a quella di un testo a forte vocazione tecnica rispetto a quanto potessero far pensare i dati sulla distribuzione del VdB.

Da ultimo vi è il dato sulla densità lessicale (0,571), che conferma quanto ipotizzato dalla letteratura a proposito di questo parametro: il testo ha un'elevata densità, il che significa che riversa maggiori informazioni, quindi maggior carico cognitivo, su chi legge, mantenendo elevato il proprio tasso di complessità.

CAUnifi-Ri (fig. 6) ha, come osserviamo, comportamenti molto simili a quelli di CAUnifi-Or, se pur con qualche modifica. In lieve aumento la percentuale di lemmi appartenente al VdB (69,3%), con la seguente distribuzione interna: F = 79,0%; AU = 18,0%; AD = 3,0%. La TTR è identica a quella di CAUnifi-Or (0,630), mentre aumenta lievemente la densità lessicale (0,582). Quest'ultimo dato potrebbe dipendere dal fatto che il testo riscritto ha, come abbiamo visto in § 3.1, un numero leggermente inferiore di *token* rispetto all'originale, dunque la densità cresce di poco. Dati i valori molto simili tra i due testi, vediamo anche in questo caso che le modifiche apportate nella riscrittura in ottica di una maggiore inclusione linguistica delle persone non binarie non sembrano, alla luce dei dati, aumentare la complessità lessicale del testo.

Procediamo ora ad analizzare il profilo lessicale di CAUnipi-Or e CaUnipi-Ri (figg. 7 e 8).

Rispetto alle oscillazioni di CAUnifi-Or, che presenta profili ora di semplicità e ora di difficoltà, ancora una volta è chiaro che CAUnipi-Or, in quanto regolamento, è eminentemente tecnico. In quest'ultimo (fig. 7), la percentuale di lemmi che fanno parte del VdB è del 63,9% (F = 76,4%; AU = 21,4%; AD = 2,1%), inferiore di

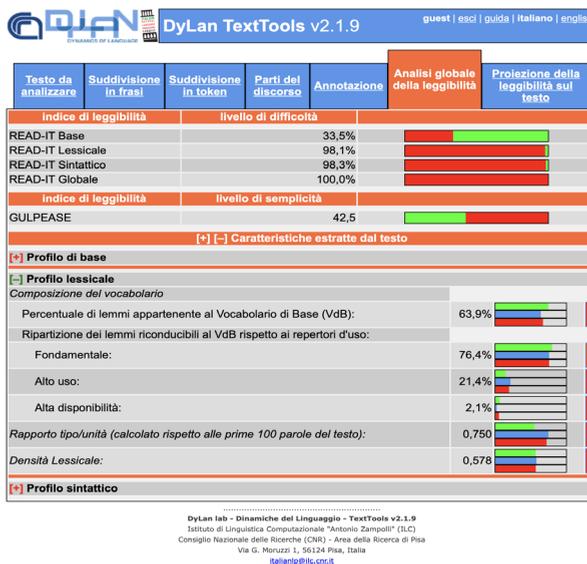


Figura 7. Profilo lessicale di CAUnipi-Or.

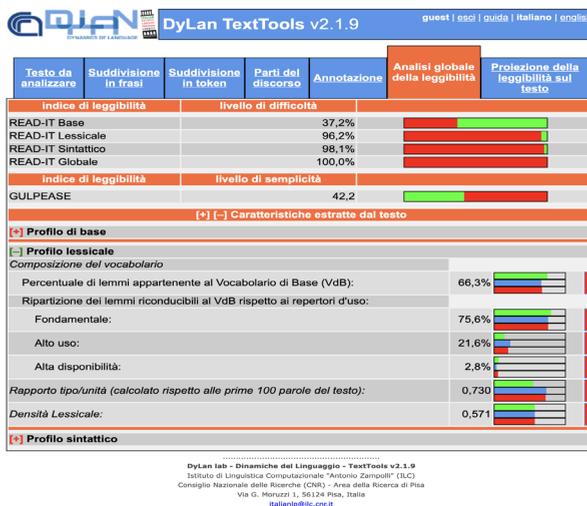


Figura 8. Profilo lessicale di CAUnipi-Ri.

diversi punti percentuali rispetto a CaUnifi-Or. In questo caso la TTR, essendo un testo tecnico, è molto elevata (0,750), sintomo del fatto che il testo è davvero lontanissimo da qualsiasi intento divulgativo. La densità lessicale si attesta su 0,578.

Per quanto riguarda CAUnipi-Ri (fig. 8), abbiamo un aumento della percentuale di lemmi appartenenti VdB rispetto all'originale, che ora rappresentano il 66,3% (F = 75,6%; AU = 21,6%; AD = 2,8%), una TTR di 0,730 - ancora alta, ma in diminuzione non significativa rispetto all'originale - e una densità lessicale lievemente più bassa del testo di partenza (0,571).

Alla luce di questi dati, possiamo dire che nessun accorgimento inclusivo sembra modificare "in meglio" il testo originale in un'ottica di una minore complessità, ma nemmeno peggiorarlo rendendolo ancora più difficile. In presenza di un testo più specialistico come un regolamento, chiaramente, le modifiche da poter fare in fase di riscrittura sono poche, pena la diminuzione della monoreferenzialità dei termini e anche il rischio di abbassamento del potere regolativo che il testo porta con sé. Tuttavia, le modifiche apportate in questo caso per un'inclusione di genere non binaria sono state minime e non hanno alterato la già complessa struttura del testo. Pertanto, pare si possa affermare anche in questo passaggio che una scrittura inclusiva non peggiorerebbe la situazione di un testo già piuttosto faticoso, che semmai avrebbe bisogno di altri accorgimenti per risultare più fruibile.

3.3 Profilo sintattico e sentence readability

In questa sezione verrà indagato il profilo sintattico dei testi originali e riscritti, mostrando poi anche le proiezioni di leggibilità delle singole frasi all'interno dei testi. Rispetto ai tradizionali strumenti di misurazione della leggibilità, READ-IT ha infatti il pregio di monitorare sia la leggibilità del testo nella sua globalità, sia quella della singola frase (Dell'Orletta, Montemagni & Venturi 2014). Come è stato ampiamente dimostrato in letteratura, la sintassi del linguaggio amministrativo è complessa e spesso debitrice nei confronti di quella giuridica, in particolare per quanto riguarda i bandi e i regolamenti¹⁸. Proprio per questo, la capillare possibilità di monitorare sul testo la fruibilità delle singole frasi può farci vedere con più chiarezza i luoghi di complessità e quelli di semplicità. Poiché CAUnifi-Or e CAUnipi-Or sono di lunghezza diver-

¹⁸ Per un approfondimento sulla sintassi della lingua giuridica cfr. Gualdo & Telve (2011: 441 ss.).

sa tra loro, ma CAUnifi-Ri e CAUnipi-Ri sono molto simili ciascuna al proprio originale, l'analisi della leggibilità per frase verrà bilanciata in questo modo: saranno mostrati i primi 20 periodi di CaUnifi-Or e CAUnifi-Ri e i primi 3 articoli, per un totale di 26 periodi, di CAUnipi-Or e CAUnipi-Ri.

L'analisi dei testi condotta con READ-IT permette di enucleare informazioni sulle varie combinazioni di tratti morfosintattici e sintattici dei testi stessi. Innanzitutto, grazie alla suddivisione in parti del discorso e all'annotazione, ci vengono restituite le percentuali delle categorie morfosintattiche presenti; in secondo luogo, vengono fatte riflessioni sulla struttura a dipendenze dell'intero testo, esplorando le articolazioni interne dei singoli periodi, il rapporto tra proposizioni principali e proposizioni subordinate, le articolazioni della singola proposizione, la profondità dell'albero sintattico e la misura delle relazioni di dipendenza.

Nelle figg. 9 e 10 possiamo vedere il profilo sintattico di CAUnifi-Or e CAUnifi-Ri

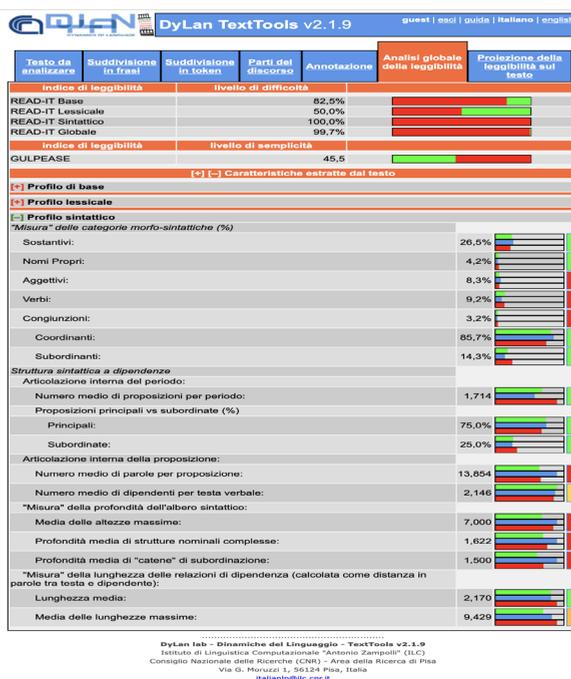
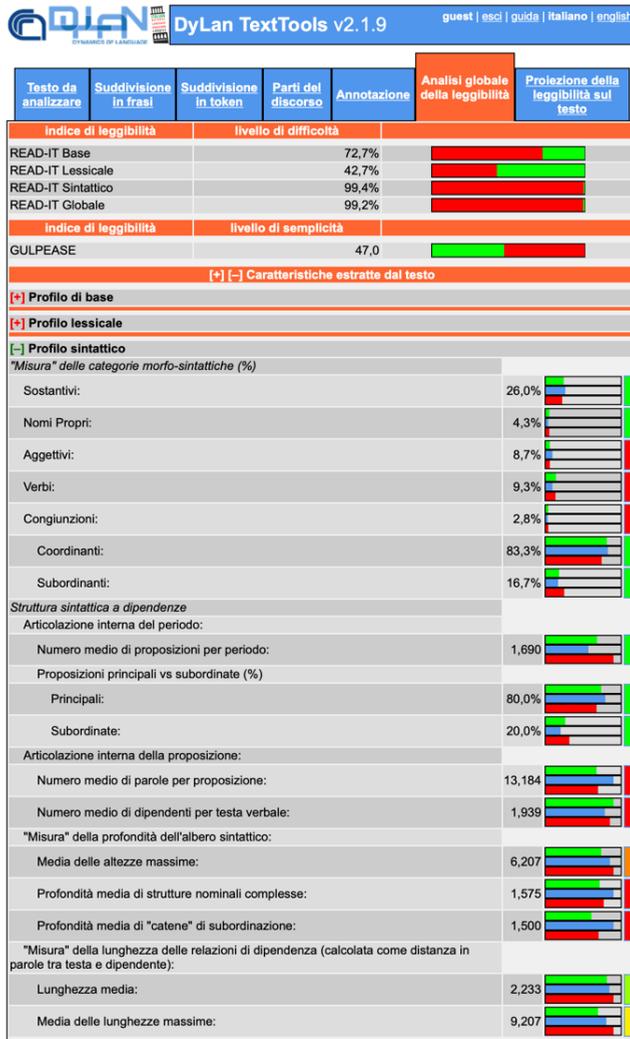


Figura 9. Profilo sintattico di CAUnifi-Or.



DyLan lab - Dinamiche del Linguaggio - TextTools v2.1.9
 Istituto di Linguistica Computazionale "Antonio Zampolli" (ILC)
 Consiglio Nazionale delle Ricerche (CNR) - Area della Ricerca di Pisa
 Via G. Moruzzi 1, 56124 Pisa, Italia
italianlp@ilc.cnr.it

Figura 10. Profilo sintattico di CAUnifi-Ri.

Iniziando da CAUnifi-Or (fig. 9), per quanto riguarda la misura delle percentuali delle categorie morfosintattiche lungo il testo, il primo elemento da notare è il rapporto tra nomi e verbi (N/V). Come spiega Montemagni (2013: 13-14), il fatto che il rapporto di frequenza N/V sia intrinsecamente legato al genere del testo e a questioni di natura diamesica e diafasica può considerarsi un dato acquisito¹⁹. Dai dati analizzati dalla stessa Montemagni (2013) risulta che i testi informativi e regolativi abbiano percentuali di N nettamente superiori alla media²⁰. Come vediamo, in CAUnifi-Or i sostantivi e i nomi propri rappresentano rispettivamente il 26,5% e il 4,2% (totale: 30,7%) degli elementi presenti, laddove i verbi rappresentano il 9,2%. Da questo punto di vista, pur trattandosi del testo di un sito *web* teoricamente pensato per la lettura più agevole, CAUnifi-Or si comporta come un testo tecnico a tutti gli effetti, con un importante carico informativo veicolato dall'elevato portato degli elementi nominali. Da sottolineare è anche la bassa presenza di aggettivi (8,3%), peculiare del genere burocratico-amministrativo, il quale è poco incline all'arricchimento dato dagli attributi (Brunato 2014). Pure le congiunzioni sono in una percentuale molto bassa (3,2%).

Dal punto di vista del numero medio di proposizioni per periodo e del rapporto tra principali e subordinate, vediamo che CAUnifi-Or è considerato di facile lettura, con un numero medio di proposizioni per periodo (1,714) persino più basso di quello dei testi di *training* con cui il sistema è stato addestrato per il livello della lettura facile, e una forte preponderanza di principali (75,0%) rispetto alle subordinate (25,0%). Questo aspetto potrebbe dipendere dal fatto che l'organizzazione del testo secondo il modello Q&A stimola la formulazione di tanti periodi diversi e separati, più che far proliferare la subordinazione. Tuttavia, l'articolazione interna alla singola proposizione è invece quella tipica di un testo complesso, con un elevato numero medio di *token* per proposizione (13,854) e una media di più di due dipendenti per testa verbale (2,146).

Il testo risulta difficile pure sulla base delle analisi della profondità dell'albero sintattico, con un'elevata media delle altezze massime (7,000), strutture

¹⁹ Nel contesto citato, Montemagni fa riferimento al lavoro di Voghera (2005), al quale si rinvia per ulteriori approfondimenti.

²⁰ Nel lavoro di Montemagni (2013) il confronto è con testi di tipo creativo.

nominali complesse piuttosto profonde (1,622) e un altrettanto elevato valore medio di profondità delle catene di subordinazione (1,500). Infine, se la lunghezza media delle relazioni di dipendenza è bassa (2,170), la media delle lunghezze massime (9,429) si attesta invece su un livello intermedio.

CAUnifi-Ri (fig. 10) mantiene valori molto simili, o presenta variazioni lievi, rispetto a CAUnifi-Or, come testimoniato anche dall'oscillazione non elevata del valore di complessità sintattica dei due testi che si vede nella panoramica generale e che già avevamo menzionato all'inizio di questo contributo (100% in CAUnifi-Or VS 99,4% in CAUnifi-Ri). Anche nel caso di quest'ultimo, i sostantivi (26,0%) e i nomi propri (4,3%) costituiscono insieme quasi un terzo dei *token* totali del testo (30,3%), mentre i verbi e gli aggettivi rappresentano rispettivamente il 9,3% e l'8,7% dei *token* e le congiunzioni solo il 2,8%: non pare dunque alterato rispetto all'originale il rapporto N/V, testimone di un testo complesso dall'elevato carico informativo veicolato dai N.

Per quanto riguarda la struttura sintattica a dipendenze, il numero medio di proposizioni per periodo è 1,690 - anche in questo caso molto più basso di quello del *training corpus* del polo semplice e persino lievemente più basso di CAUnifi-Or, in diminuzione non significativa - mentre rispetto a quest'ultimo in CAUnifi-Ri cresce ulteriormente il rapporto tra principali e subordinate a favore delle prime (80,0% VS 20,0 %). Il numero di parole per proposizione rimane comunque elevato (13,184), anche se più basso che in CAUnifi-Or, come pure lievemente più basso è il numero medio di dipendenti per testa verbale (1,939). Per quanto riguarda le misure della profondità dell'albero sintattico, in CAUnifi-Ri calano lievemente la media delle altezze massime (6,207) e la profondità media delle strutture nominali complesse (1,575) rispetto a CAUnifi-Or, mentre rimane identico il valore della profondità media delle catene di subordinazione (1,500). Infine, dal punto di vista della lunghezza delle relazioni di dipendenza abbiamo una media di 2,233, leggermente più alta di quella di CAUnifi-Or, mentre la media delle lunghezze massime è lievemente più bassa (9,207) dell'originale. Nonostante queste modeste oscillazioni, possiamo tuttavia sostenere che la riscrittura con accorgimenti inclusivi non binari non abbia semplificato il testo rendendolo meno tecnico e più fruibile, tuttavia non lo abbia neppure complicato.

Le seguenti figg. 11 e 12 mostrano la proiezione di leggibilità per singola frase sia di CAUnifi-Or che di CAUnifi-Ri.

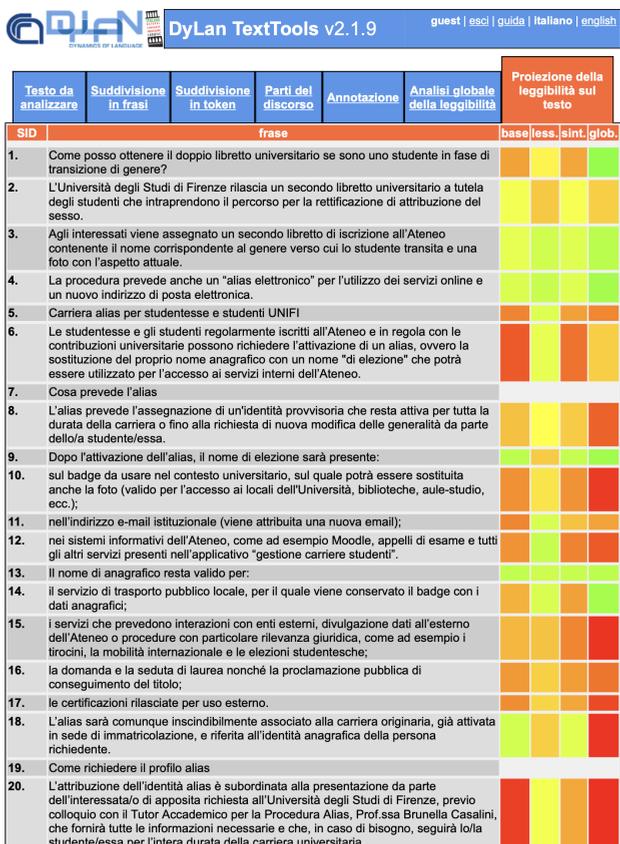


Figura 11. Sentence readability dei primi 20 periodi di CAUnif-Or.

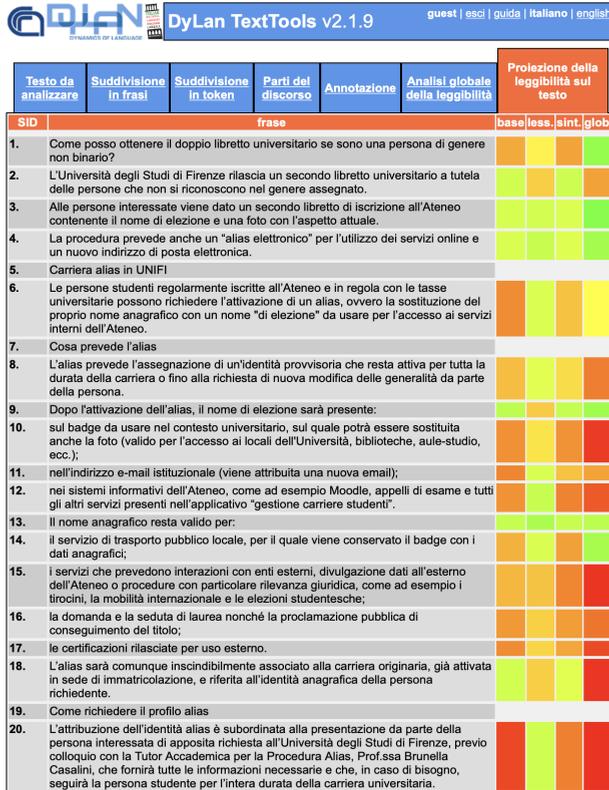


Figura 12. Sentence readability dei primi 20 periodi di CAUnifi-Ri.

Come possiamo evincere dalle figure, la leggibilità delle singole frasi sul testo rimane pressoché identica, se non per qualche minima oscillazione in alcuni passaggi. In media, in entrambi i testi nella prima parte (frasi 1-4) la lettura globale è ritenuta facile o di media difficoltà, andando a complicarsi nella parte centrale, che è anche quella dei periodi più articolati. Anche in questo caso, i dati sembrano confermare l'ipotesi che una riscrittura e attenta alle pluralità di genere non pregiudichi la leggibilità del testo, che semmai risulta già compromessa per altri motivi che non hanno a che vedere con le tematiche di genere.

Operiamo ora le analisi sintattiche su CAUnipi-Or e CAUnipi-Ri (figg. 13 e 14).

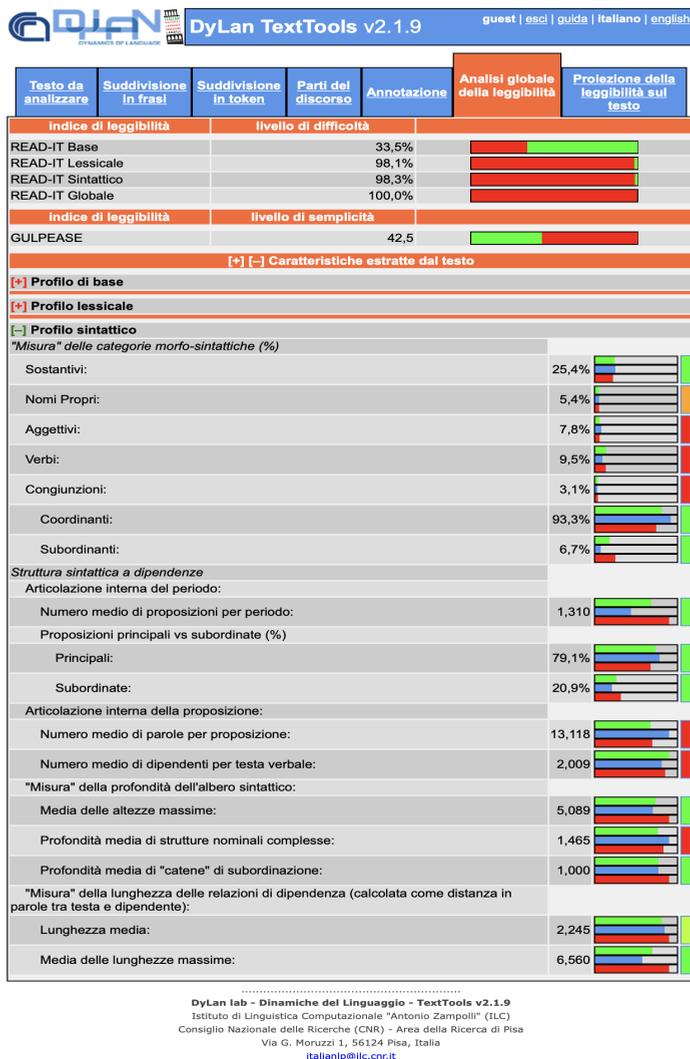
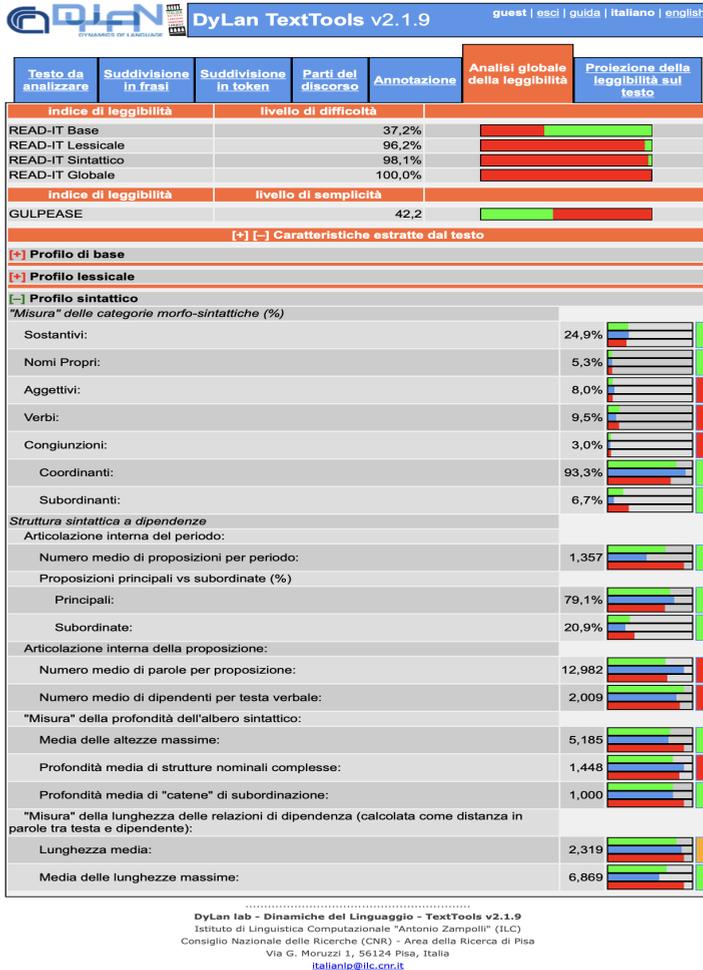


Figura 13. Profilo sintattico di CAUnipi-Or.

Figura 14. *Profilo sintattico di CAUnipi-Ri.*

Iniziando da CAUnipi-Or (fig. 13), notiamo che i sostantivi (25,4%) e i nomi propri (5,4%) rappresentano il 30,8% dei *token* del testo, mentre i verbi il 9,5% e gli aggettivi il 7,8%: anche in questo caso, il rapporto N/V è molto sbilanciato a favore dei primi - più del triplo dei secondi - e gli aggettivi sono sot-

to-rappresentati (7,8%), proprio come ci si aspetterebbe da un testo tecnico; sotto-rappresentate sono pure le congiunzioni (3,1%).

Nell'ambito della struttura sintattica a dipendenze, notiamo che il numero medio di proposizioni per periodo (1,310) è sensibilmente più basso della soglia del *training corpus* per il polo della facile lettura. Ciò dipende probabilmente dal fatto che in un regolamento viene adottato il cosiddetto 'stile commatico' "caratterizzato da una minuta suddivisione del testo (titoli, capi, articoli, commi, sottocommi ecc.), che a sua volta è segmentato in periodi brevissimi seguiti da punto a capo [...], e da una rigorosa messa in rilievo dei punti nodali del testo" (Gualdo & Telve 2011: 448); notiamo che anche la percentuale di proposizioni principali (79,1%) rispetto alle subordinate (20,9%) è molto elevata.

Come già accaduto in precedenza per CAUnifi-Or, anche in questo caso a un'articolazione interna del periodo piuttosto semplice, che fa illuminare la luce verde del *software*, corrisponde un'articolazione interna della proposizione che invece è complessa: per ognuna di esse ci sono mediamente 13,118 parole, e il numero medio di dipendenti per testa verbale è 2,009. L'albero sintattico presenta una bassa media delle altezze massime (5,089), con una profondità media delle strutture nominali complesse di 1,465 e una delle catene di subordinazione di 1,000.

Infine, per quanto riguarda le relazioni di dipendenza, abbiamo una lunghezza media di 2,245 e una media delle lunghezze massime di 6,560. Contrariamente a quanto ci possiamo aspettare da un testo pensato per essere più tecnico come un regolamento, molti di questi dati sono invece più vicini ai valori espressi dal *training corpus* di semplice lettura rispetto a quelli del *training corpus* di lettura complessa, come difatti si evince dalle luci verdi accanto al singolo parametro.

Passando ora CAUnipi-Ri (fig. 14), vediamo che anche in questo caso esso presenta valori quasi identici, tranne qualche minima oscillazione, rispetto a quelli di CAUnipi-Or. Il rapporto N/V è ancora fortemente sbilanciato verso i primi (24,9% di sostantivi e 5,3% di nomi propri, per un totale di 30,2%, a fronte del 9,5% dei verbi, valore che rimane identico rispetto a CAUnipi-Or), suggerendo che ci troviamo di fronte a un testo costruito con un notevole carico informativo veicolato dagli elementi nominali. Gli aggettivi (8,0%), pur con una oscillazione non significativa verso l'alto rispetto a CAUnipi-Or, continuano a essere sotto-rappresentati, come le congiunzioni (3,0%).

Il basso numero medio di proposizioni per periodo (1,357) e l'importante preponderanza di proposizioni principali rispetto alle subordinate (79,1% di principali VS 20,9% di subordinate, valori identici a quelli di CAUnipi-Or) suggeriscono un'articolazione interna del periodo in CAUnipi-Ri orientata al polo semplice. Tuttavia, l'articolazione interna della proposizione mantiene, come nell'originale, un importante livello di difficoltà, con un numero medio di parole per proposizione di 12,982 (il lieve flessione rispetto a CAUnipi-Or, ma comunque elevato) e un numero medio di dipendenti per testa verbale di 2,009, di nuovo uguale a CAUnipi-Or. A proposito delle misure della profondità dell'albero sintattico, notiamo che la media delle altezze massime è 5,185, la profondità media delle strutture nominali complesse è 1,448 (entrambi i valori in aumento non significativo rispetto CAUnipi-Or) e infine la profondità media delle catene di subordinazione è 1,000, identica al testo di partenza. Da ultimo, per quanto riguarda le relazioni di dipendenza, notiamo che la loro lunghezza media è di 2,319, mentre la media delle lunghezze massime è di 6,869, entrambe in aumento lieve rispetto CaUnipi-Or, non tale da attribuire a CaUnipi-Ri una complicazione testuale maggiore rispetto all'originale.

In ultima battuta, nelle figg. 15 e 16 possiamo osservare i profili di *sentence readability* dei primi tre articoli di CAUnipi-Or e CaUnipi-Ri.

 DyLan TextTools v2.1.9 quest esci guida italiano english		Protezione della leggibilità sul testo							
Testo da analizzare	Suddivisione in frasi	Suddivisione in token	Parti del discorso	Annotazione	Analisi globale della leggibilità	base	less.	sint.	glob.
SID	frase					base	less.	sint.	glob.
1.	Regolamento per l'attivazione e la gestione delle carriere alias								
2.	D.R. n. 177/2020 del 30 gennaio 2020 - Emanazione								
3.	Articolo 1 Finalità								
4.	1.								
5.	La finalità del presente Regolamento è garantire il benessere psico-fisico degli/le studenti e del personale di Ateneo, a qualsiasi titolo facente parte della comunità universitaria pisana, nonché favorire la realizzazione di un ambiente di studio e di lavoro inclusivo, sempre ispirato al valore fondante della pari dignità delle persone, eliminando situazioni di disagio per coloro che intendono modificare nome e identità nell'espressione della propria autodeterminazione di genere.								
6.	2.								
7.	L'Università di Pisa, nel perseguimento degli obiettivi che ispirano il Comitato unico di garanzia "per le pari opportunità, la valorizzazione del benessere di chi lavora e contro le discriminazioni" (di seguito "CUG"), nel rispetto della normativa vigente, pone in essere le misure di protezione per le persone che abbiano la necessità di utilizzare, all'interno dell'Ateneo, un nome diverso rispetto a quello anagrafico, mediante l'attivazione di una carriera alias.								
8.	3.								
9.	L'Università di Pisa, in accordo e attraverso le azioni positive del CUG, con l'obiettivo di garantire il pieno e positivo accesso al dispositivo della carriera alias, si impegna a fornire adeguata formazione in merito ai temi che riguardano l'identità di genere a tutto il personale coinvolto nelle procedure relative alla carriera alias e al personale in contatto diretto con i soggetti richiedenti.								
10.	Articolo 2 Destinatari/e								
11.	1.								
12.	L'attivazione di una carriera alias può essere richiesta da tutti/e i/e componenti della comunità universitaria, docenti, studenti, personale tecnico-amministrativo, dirigenti, componenti esterni/e degli organi collegiali e quanti/e a vario titolo operano, anche occasionalmente e temporaneamente, nelle strutture dell'Ateneo.								
13.	2.								
14.	Il/La delegato/a alle attività "Gender Studies and Equal Opportunities" può autorizzare il riconoscimento temporaneo di una identità alias in favore di ospiti dell'Ateneo che, in occasione di attività seminari, convegni, cicli di lezioni o altre iniziative, intendano utilizzare su locandine, badge, ecc. un nome alias scelto in sostituzione dell'identità anagrafica ufficiale.								
15.	Articolo 3								
16.	Procedura per l'attivazione carriera alias								
17.	1.								
18.	L'attivazione della carriera alias avviene mediante presentazione al/la delegato/a alle attività "Gender Studies and Equal Opportunities" dell'istanza redatta secondo il modello allegato (All.to 1), corredata da un documento di identità.								
19.	2.								
20.	Il/La delegato/a alle attività "Gender Studies and Equal Opportunities" assume le funzioni di tutor accademico per la carriera alias, supporta la persona richiedente nella procedura di richiesta e nella successiva gestione della carriera alias.								
21.	3.								
22.	Il/La delegato/a, valutata l'istanza, acquisito il parere del/la Responsabile della Protezione Dati (di seguito "RPD") dell'Ateneo (RPDcarrieraalias@unipi.it), può autorizzare l'attivazione della carriera alias.								
23.	4.								
24.	L'attivazione della carriera è subordinata alla sottoscrizione, da parte del/la Rettore/ric e del/la Richiedente, di un Accordo di riservatezza, redatto secondo i modelli allegati (All.ti 2, 3 e 4).								
25.	5.								
26.	All'Accordo di riservatezza fa seguito un provvedimento del/la RPD contenente le istruzioni operative per gli uffici competenti.								

Figura 15. *Sentence readability dei primi tre articoli di CaUnipi-Or.*

Testo da analizzare		Suddivisione in frasi	Suddivisione in token	Parti del discorso	Annotazione	Analisi globale della leggibilità	Proiezione della leggibilità sul testo			
SID	frase						base	less.	sint.	glob.
1.	Regolamento per l'attivazione e la gestione delle carriere alias									
2.	D.R. n. 177/2020 del 30 gennaio 2020 - Emanazione									
3.	Articolo 1 Finalità									
4.	1.									
5.	La finalità del presente Regolamento è garantire il benessere psico-fisico delle persone studenti e del personale di Ateneo, a qualsiasi titolo facente parte della comunità universitaria pisana, nonché favorire la realizzazione di un ambiente di studio e di lavoro inclusivo, sempre ispirato al valore fondante della pari dignità delle persone, eliminando situazioni di disagio per coloro che intendono modificare nome e identità nell'espressione della propria autodeterminazione di genere.									
6.	2.									
7.	L'Università di Pisa, nel perseguimento degli obiettivi che ispirano il Comitato unico di garanzia "per le pari opportunità, la valorizzazione del benessere di chi lavora e contro le discriminazioni" (di seguito "CUG"), nel rispetto della normativa vigente, pone in essere le misure di protezione per le persone che abbiano la necessità di utilizzare, all'interno dell'Ateneo, un nome diverso rispetto a quello anagrafico, mediante l'attivazione di una carriera alias.									
8.	3.									
9.	L'Università di Pisa, in accordo e attraverso le azioni positive del CUG, con l'obiettivo di garantire il pieno e positivo accesso al dispositivo della carriera alias, si impegna a fornire adeguata formazione in merito ai temi che riguardano l'identità di genere a tutto il personale coinvolto nelle procedure relative alla carriera alias e al personale in contatto diretto con i soggetti richiedenti.									
10.	Articolo 2 Persone destinatarie									
11.	1.									
12.	L'attivazione di una carriera alias può essere richiesta da tutte le persone che compongono la comunità universitaria, ovvero persone docenti, persone studenti, personale tecnico-amministrativo, persone dirigenti, figure componenti esterne degli organi collegiali e da chiunque a vario titolo operi, anche occasionalmente e temporaneamente, nelle strutture dell'Ateneo.									
13.	2.									
14.	La persona delegata alle attività "Gender Studies and Equal Opportunities" può autorizzare il riconoscimento temporaneo di una identità alias in favore di persone ospiti dell'Ateneo che, in occasione di attività seminari, convegni, cicli di lezioni o altre iniziative, intendano utilizzare su locandine, badge, ecc. un nome alias scelto in sostituzione dell'identità anagrafica ufficiale.									
15.	Articolo 3 Procedura per l'attivazione carriera alias									
16.	1.									
17.	L'attivazione della carriera alias avviene mediante presentazione alla persona delegata alle attività "Gender Studies and Equal Opportunities" dell'istanza redatta secondo il modello allegato (All.to 1), corredata da un documento di identità.									
18.	2.									
19.	La persona delegata alle attività "Gender Studies and Equal Opportunities" assume le funzioni di tutor d'ateneo per la carriera alias, supporta la persona richiedente nella procedura di richiesta e nella successiva gestione della carriera alias.									
20.	3.									
21.	La persona delegata, valutata l'istanza, acquisito il parere della persona Responsabile della Protezione Dati (di seguito "persona RPD") dell'Ateneo (RPDcarrieraalias@unipi.it), può autorizzare l'attivazione della carriera alias.									
22.	4.									
23.	L'attivazione della carriera è subordinata alla sottoscrizione, da parte del Rettore/della Rettrice e della persona richiedente, di un Accordo di riservatezza, redatto secondo i modelli allegati (All.ti 2, 3 e 4).									
24.	5.									
25.	All'Accordo di riservatezza fa seguito un provvedimento della persona RPD contenente le istruzioni operative per gli uffici competenti.									

Figura 16. Sentence readability dei primi tre articoli di CaUnipi-Ri.

Anche in questo caso, vediamo come, nonostante alcuni rari luoghi di semplicità, la leggibilità per frasi di entrambi i testi è piuttosto scarsa e si mantiene più o meno identica tanto nella versione originale quanto in quella riscritta.

Alla luce delle riflessioni esperite fino a questo momento sui dati analizzati, pare che si possano fare alcune considerazioni.

Innanzitutto, come predetto dalla letteratura, possiamo sostenere che alcuni profili di semplicità sintattica possano ravvisarsi anche in testi a elevato tasso di complessità, e ciò può dipendere da diversi fattori; tuttavia, e ciò emerge dallo studio, questo non ne altera il livello di leggibilità, che risulta da una valutazione globale e rimane comunque, in media, piuttosto basso.

In secondo luogo, questione più pregnante per questo contributo, notiamo che questi testi hanno un importante livello di complessità generale che ne rende molto faticosa la fruizione, e che ciò non peggiora - anzi, qualche volta si apprezzano persino lievi miglioramenti - all'adozione di strategie linguistiche rappresentative di identità di genere non solo binarie. L'obiezione, talvolta mossa, che l'adozione di un linguaggio rispettoso delle pluralità di genere renda i testi più complessi non trova perciò riscontro nei dati qui analizzati; di conseguenza, sempre rimanendo sui dati analizzati in questa sede, non si può usare l'aumento della complessità dei testi come argomento a supporto per scoraggiare l'adozione di un linguaggio più inclusivo nell'ambito della comunicazione istituzionale.

4. Verso il futuro: il corpus ALIAS

I dati analizzati finora ci suggeriscono che, a fronte di strategie non solo binarie, la leggibilità della comunicazione istituzionale di questo dominio non acquisisce maggiore complessità di quella che già la caratterizza. Tuttavia, i tentativi di riscrittura qui operati sono stati, per ovvi motivi, estremamente parziali, poiché hanno riguardato i testi di due sole università presi a modello: ne consegue che non siamo in grado, al momento, di trarre da questi dati assunzioni di respiro generale.

Per tale motivo, oltre all'analisi dei dati, il presente contributo vuole anche offrire un'idea per future applicazioni volte a migliorare la comunicazione istituzionale nel dominio delle carriere *alias*.

Il punto di partenza sarebbe la ripresa dei principi che hanno animato il cosiddetto *Repertorio COINFO* (Consorzio interuniversitario per la formazione), ovvero "la raccolta di modelli di comunicazione standard delle università" (Cortelazzo 2021: 134). Come spiega lo studioso (*ibidem*), infatti,

le istituzioni dello stesso tipo si trovano tutte nella necessità di far conoscere al proprio pubblico di riferimento la stessa informazione [...]; ognuna, però,

provvede da sé a scrivere queste informazioni, che, al di là delle specificità dovute all'autonomia decisionale, hanno un nucleo comune. Si ha così un enorme spreco di risorse: un considerevole numero di dipendenti, con gradi diversi di abilità di scrittura, si trova a dover scrivere lo stesso testo, o testi tra loro molto simili. Ne consegue che è alto il rischio che molti dei prodotti di questa così diffusa attività di scrittura risultino inefficaci.

Ecco, allora, che può essere utile pubblicare raccolte di testi standard, da mettere a disposizione di tutti gli enti che hanno bisogno di diffondere lo stesso tipo di informazioni: il lavoro di pianificazione del testo [...] viene fatto una volta per tutte da un gruppo di scriventi che può dedicare a questa fase cruciale tutto il tempo necessario. Poi, i dipendenti di ogni università possono copiare e adattare alle proprie necessità i modelli presentati, partendo da una fonte redatta con attenzione [...].

In questo passaggio Cortelazzo fa riferimento ai suggerimenti per la semplificazione della comunicazione istituzionale e non – o non necessariamente – all'inclusione di genere, ma si potrebbe applicare lo stesso principio ai testi per le carriere *alias*. Tutti gli atenei che offrono questo servizio hanno infatti almeno un regolamento di ateneo approvato con un Decreto Rettorale, almeno un modulo (spesso chiamato *Accordo di riservatezza*) da compilare e firmare, e almeno una pagina di sito *web* dedicata alle più sommarie informazioni su questo argomento; molti atenei hanno anche una pagina all'interno del portale di ogni persona, dalla quale si può accedere alla richiesta del servizio.

L'idea che qui si vuole suggerire è quella di coniugare il possibile meccanismo di funzionamento del Repertorio COINFO con gli strumenti linguistico-computazionali a disposizione per l'analisi dei testi, creando un *corpus*, che chiameremo *ALIAS*. Tale *corpus* dovrà comprendere tutti i testi riguardanti le carriere *alias* prodotti da tutti gli atenei italiani, pubblici e privati. Eventualmente esso potrà essere ripartito in sotto-*corpora* che suddividono le diverse tipologie testuali: il regolamento, il modulo, la pagina informativa di sito *web*, la pagina riservata del portale, etc.

In una prima fase, i testi originali dovranno essere sottoposti al monitoraggio della leggibilità tramite READ-IT: ciò consentirà di avere una panoramica generale dello "stato di salute" dei testi istituzionali universitari per le carriere *alias*, panoramica che, rispetto allo stato attuale, sarà basata su una maggiore quantità di dati e quindi rifletterà più fedelmente la realtà.

Dopodiché, grazie ai risultati offerti da questa prima fase di verifica, si potranno operare delle riscritture analoghe a quelle qui proposte nell'Appendice. Esse, questa volta, dovranno esplicitamente tentare di integrare le strategie di inclusione non binaria con le istanze per la comunicazione istituzionale chiara e comprensibile. Le riscritture dovranno tener conto dei punti di forza e dei punti di debolezza emersi in fase di monitoraggio per ogni tipologia testuale, al fine di arrivare a produrre, per ogni tipologia, un modello di testo standard che possa funzionare, con pochissimi adattamenti, per il maggior numero possibile di atenei.

La speranza è che un lavoro simile sia in grado di fornire, traendo le caratteristiche più funzionali dai vari testi reali, modelli di testi allo stesso tempo chiari, comprensibili e inclusivi, per renderli disponibili e adottabili dalle istituzioni universitarie su ampia scala.

5. Conclusioni

In questo contributo si è cercato di dare una risposta ai quesiti dai quali abbiamo preso le mosse, ovvero se sia possibile proporre una comunicazione istituzionale rappresentativa di identità di genere non più solo binarie, se questi accorgimenti possano entrare in qualche modo in rapporto con quelli per la comunicazione chiara e comprensibile, e infine come e in quale percentuale gli strumenti di Trattamento Automatico del Linguaggio possano essere d'aiuto per questi scopi. Le domande sembrano aver avuto, almeno limitatamente ai risultati emersi lungo l'articolo, una risposta positiva. Un *tool* come READ-IT è particolarmente utile per monitorare la condizione attuale della comunicazione istituzionale per le carriere *alias* e il suo livello di chiarezza, come già l'ampia letteratura in merito ci aveva dimostrato da tempo. In parallelo, si possono proporre riscritture che tengano conto dell'inclusione di genere non binaria, molto importante dato che ci troviamo in una branca della comunicazione istituzionale esplicitamente dedicata alla tutela degli interessi e della dignità delle persone *gender variant*.

In una prospettiva futura più ampia, si potranno raggiungere risultati molto più concreti e basati su una maggiore quantità di dati tramite l'applicazione delle metodologie quantitative ai testi del nostro dominio d'interesse. In questo modo, auspicabilmente, si risponderà alla sfida di fronte

alla quale l'università contemporanea ci pone, ossia quella di proporre testi che siano certamente comprensibili, chiari ed efficaci secondo le più comuni linee guida per la scrittura istituzionale, ma che allo stesso tempo tengano in considerazione l'importanza di includere tutte le identità di genere presenti nelle comunità accademiche.

Appendice

Testo tratto dal sito *web* dell'Università di Firenze: versione originale e versione riscritta

Come posso ottenere il doppio libretto universitario se sono uno studente in fase di transizione di genere?

L'Università degli Studi di Firenze rilascia un secondo libretto universitario a tutela degli studenti che intraprendono il percorso per la rettificazione di attribuzione del sesso. Agli interessati viene assegnato un secondo libretto di iscrizione all'Ateneo contenente il nome corrispondente al genere verso cui lo studente transita e una foto con l'aspetto attuale.

La procedura prevede anche un "alias elettronico" per l'utilizzo dei servizi online e un nuovo indirizzo di posta elettronica.

Carriera alias per studentesse e studenti UNIFI

Le studentesse e gli studenti regolarmente iscritti all'Ateneo e in regola con le contribuzioni universitarie possono richiedere l'attivazione di un alias, ovvero la sostituzione del proprio nome anagrafico con un nome "di elezione" che potrà essere utilizzato per l'accesso ai servizi interni dell'Ateneo.

Cosa prevede l'alias

L'alias prevede l'assegnazione di un'identità provvisoria che resta attiva per tutta la durata della carriera o fino alla richiesta di nuova modifica delle generalità da parte dello/a studente/essa.

Come posso ottenere il doppio libretto universitario se **sono una persona di genere non binario**?

L'Università degli Studi di Firenze rilascia un secondo libretto universitario a tutela delle **persone che non si riconoscono nel genere assegnato. Alle persone interessate** viene dato un secondo libretto di iscrizione all'Ateneo contenente **il nome di elezione** e una foto con l'aspetto attuale²¹.

La procedura prevede anche un "alias elettronico" per l'utilizzo dei servizi online e un nuovo indirizzo di posta elettronica.

Carriera alias **in** UNIFI

Le **persone studenti** regolarmente iscritte all'Ateneo e in regola con le tasse universitarie possono richiedere l'attivazione di un alias, ovvero la sostituzione del proprio nome anagrafico con un nome **eletto** che potrà essere utilizzato per l'accesso ai servizi interni dell'Ateneo.

Cosa prevede l'alias

L'alias prevede l'assegnazione di un'identità provvisoria che resta attiva per tutta la durata della carriera o fino alla richiesta di nuova modifica delle generalità da parte **della persona studente**.

²¹ In questo caso, la modifica ha riguardato il contenuto oltre che la forma. Infatti, non solo le persone che decidono di intraprendere un percorso ufficiale di rettifica dei documenti d'identità dovrebbero poter accedere alla carriera alias, dato che essa non ha valore legale al di fuori delle mura dell'ateneo. Essa dovrebbe invece essere accessibile anche tutti quegli individui che al momento non si riconoscono nel genere assegnato loro alla nascita, a prescindere da quale sarà poi la loro decisione sulle operazioni chirurgiche, su percorsi medicalizzati in generale, oppure sulla modifica dei documenti.

Dopo l'attivazione dell'alias, il nome di elezione sarà presente:

- sul badge da usare nel contesto universitario, sul quale potrà essere sostituita anche la foto (valido per l'accesso ai locali dell'Università, biblioteche, aule-studio, ecc.);
- nell'indirizzo e-mail istituzionale (viene attribuita una nuova email);
- nei sistemi informativi dell'Ateneo, come ad esempio Moodle, appelli di esame e tutti gli altri servizi presenti nell'applicativo "Gestione carriere studenti".

Il nome di anagrafico resta valido per:

- il servizio di trasporto pubblico locale, per il quale viene conservato il badge con i dati anagrafici;
- i servizi che prevedono interazioni con enti esterni, divulgazione dati all'esterno dell'Ateneo o procedure con particolare rilevanza giuridica, come ad esempio i tirocini, la mobilità internazionale e le elezioni studentesche;
- la domanda e la seduta di laurea nonché la proclamazione pubblica di conseguimento del titolo;
- le certificazioni rilasciate per uso esterno. L'alias sarà comunque inscindibilmente associato alla carriera originaria, già attivata in sede di immatricolazione, e riferita all'identità anagrafica della persona richiedente

Dopo l'attivazione dell'alias, il nome di elezione sarà presente:

- sul badge da usare nel contesto universitario, sul quale potrà essere sostituita anche la foto (valido per l'accesso ai locali dell'Università, biblioteche, aule-studio, ecc.);
- nell'indirizzo e-mail istituzionale (viene attribuita una nuova email);
- nei sistemi informativi dell'Ateneo, come ad esempio Moodle, appelli di esame e tutti gli altri servizi presenti nell'applicativo "**Gestione carriere studenti**"²².

Il nome anagrafico resta valido per:

- il servizio di trasporto pubblico locale, per il quale viene conservato il badge con i dati anagrafici;
- i servizi che prevedono interazioni con enti esterni, divulgazione dati all'esterno dell'Ateneo o procedure con particolare rilevanza giuridica, come ad esempio i tirocini, la mobilità internazionale e le elezioni studentesche;
- la domanda e la seduta di laurea nonché la proclamazione pubblica di conseguimento del titolo;
- le certificazioni rilasciate per uso esterno. L'alias sarà comunque inscindibilmente associato alla carriera originaria, già attivata in sede di immatricolazione, e riferita all'identità anagrafica della persona richiedente.

²² In questo caso non è stato possibile operare una riscrittura eliminando il maschile sovraesteso in quanto si tratta del nome dell'applicativo, che eventualmente andrebbe cambiato con una decisione di ateneo, ma non può essere modificato in una fase di riscrittura testuale come questa.

Come richiedere il profilo alias

L'attribuzione dell'identità alias è subordinata alla presentazione da parte dell'interessata/o di apposita richiesta all'Università degli Studi di Firenze, previo colloquio con il Tutor Accademico per la Procedura Alias, Prof.ssa Brunella Casalini, che fornirà tutte le informazioni necessarie e che, in caso di bisogno, seguirà lo/la studente/essa per l'intera durata della carriera universitaria. Referente amministrativo è la dott.ssa Floriana Fabbri (reperibile all'indirizzo email unifiinclude.genere(AT)unifi.it) che metterà in contatto con il Tutor accademico per il colloquio. Al termine dell'incontro, per avviare l'alias è necessario sottoscrivere un accordo di riservatezza (pdf), nel quale vengono definite le condizioni d'impiego dell'identità alias all'interno dell'Università di Firenze, con particolare riferimento agli obblighi delle parti e alle opportunità. Non è necessaria alcuna documentazione per la richiesta di colloquio né per la sottoscrizione dell'accordo.

Chiusura del profilo alias

La carriera alias coesisterà con quella originaria finché perdurerà la carriera effettiva, a meno che non intervenga:

Come richiedere il profilo alias

L'attribuzione dell'identità alias è subordinata alla presentazione da parte **della persona interessata** di apposita richiesta all'Università degli Studi di Firenze, previo colloquio con **la Tutor Accademica**²³ per la Procedura Alias, Prof.ssa Brunella Casalini, che fornirà tutte le informazioni necessarie e che, in caso di bisogno, seguirà **la persona studente** per l'intera durata della carriera universitaria. **Referente amministrativa** è la dott.ssa Floriana Fabbri (reperibile all'indirizzo email unifiinclude.genere(AT)unifi.it) che metterà in contatto con la Tutor accademica per il colloquio²⁴.

Al termine dell'incontro, per avviare l'alias è necessario sottoscrivere un accordo di riservatezza (pdf), nel quale vengono definite le condizioni d'impiego dell'identità alias all'interno dell'Università di Firenze, con particolare riferimento agli obblighi delle parti e alle opportunità. Non è necessaria alcuna documentazione per la richiesta di colloquio né per la sottoscrizione dell'accordo.

Chiusura del profilo alias

La carriera alias coesisterà con quella originaria finché perdurerà la carriera effettiva, a meno che non intervenga:

²³ Sia per la figura della Tutor Accademica che per quella, menzionata poco più avanti, della Referente amministrativa, non risultano, a oggi, coinvolte persone che non si riconoscono nel binarismo di genere. Allo stesso tempo, però, sia la prof.ssa Brunella Casalini che la dott.ssa Floriana Fabbri risultano essere referenze extralinguistiche alle quali si può applicare il genere grammaticale femminile, anziché il maschile sovraesteso, che in questo caso non è rappresentativo della reale situazione. Pertanto, in questo passaggio e in tutte le successive menzioni di queste figure sono state adottate strategie morfosintattiche di espressione del femminile.

²⁴ In questo caso la riscrittura è stata operata per eliminare il maschile sovraesteso e adeguare la lingua alle referenze reali.

- una richiesta di interruzione da parte della studentessa o dello studente, o
 - una sentenza definitiva di rettifica di attribuzione del sesso, emessa dall'Autorità giudiziaria competente (solo per coloro che hanno intrapreso un percorso di transizione).
- NB: Qualora l'emissione della sentenza di cambio di genere avvenga in data successiva alla laurea, l'interessato/a potrà richiedere di registrare la modifica per consentire l'emissione di certificati allineati alle nuove generalità e una nuova pergamena (previo ritiro della vecchia).

Contatti
 E-mail: unifiinclude.genere@unifi.it
 Tutor accademico: Brunella Casalini
 Referente amministrativo: Floriana Fabbri

- una richiesta di interruzione **da parte della persona studente**, o
- **una sentenza definitiva di rettifica del genere anagrafico, emessa dall'Autorità giudiziaria competente**²⁵.

NB: Qualora l'emissione della sentenza di cambio di genere avvenga in data successiva alla laurea, **la persona interessata** potrà richiedere di registrare la modifica per consentire l'emissione di certificati allineati alle nuove generalità e una nuova pergamena (previo ritiro della vecchia).

Contatti
 E-mail: unifiinclude.genere@unifi.it
Tutor accademica: Brunella Casalini
Referente amministrativa: Floriana Fabbri

²⁵ Anche in questo caso, la riscrittura è contenutistica ed è volta a fornire un'informazione più corretta di quella data dal testo originale. Nei percorsi di riaffermazione di genere, infatti, le operazioni chirurgiche di rettifica del sesso biologico non sono obbligatorie per la richiesta di rettifica di genere anagrafico e modifica dei documenti. Non è pertanto obbligatorio trovarsi in un percorso medicalizzato di transizione che esiterà necessariamente con un intervento chirurgico: è possibile affrontare un percorso psicologico e terapeutico di affermazione di genere che consentirà al Tribunale competente di autorizzare la rettifica del genere anagrafico – di conseguenza dei documenti – anche senza chirurgia (cfr. https://www.infotrans.it/it-schede-39-rettifica_genere_italia, ultima consultazione: settembre 2023).

**Regolamento per le carriere alias approvato dall'Università di Pisa
con Decreto Rettorale il 30 gennaio 2020: versione originale
e versione riscritta**

Regolamento per l'attivazione e la gestione delle carriere alias

D.R. n. 177/2020 del 30 gennaio 2020 - Emanazione

Articolo 1 - Finalità

1. La finalità del presente Regolamento è garantire il benessere psico-fisico degli/le studenti e del personale di Ateneo, a qualsiasi titolo facente parte della comunità universitaria pisana, nonché favorire la realizzazione di un ambiente di studio e di lavoro inclusivo, sempre ispirato al valore fondante della pari dignità delle persone, eliminando situazioni di disagio per coloro che intendono modificare nome e identità nell'espressione della propria autodeterminazione di genere.

2. L'Università di Pisa, nel perseguimento degli obiettivi che ispirano il Comitato unico di garanzia "per le pari opportunità, la valorizzazione del benessere di chi lavora e contro le discriminazioni" (di seguito "CUG"), nel rispetto della normativa vigente, pone in essere le misure di protezione per le persone che abbiano la necessità di utilizzare, all'interno dell'Ateneo, un nome diverso rispetto a quello anagrafico, mediante l'attivazione di una carriera alias.

3. L'Università di Pisa, in accordo e attraverso le azioni positive del CUG, con l'obiettivo di garantire il pieno e positivo accesso al dispositivo della carriera alias, si impegna a fornire adeguata formazione in merito ai temi che riguardano l'identità di genere a tutto il personale coinvolto nelle procedure relative alla carriera alias e al personale in contatto diretto con i soggetti richiedenti.

Regolamento per l'attivazione e la gestione delle carriere alias

D.R. n. 177/2020 del 30 gennaio 2020 - Emanazione

Articolo 1 - Finalità

1. La finalità del presente Regolamento è garantire il benessere psico-fisico **delle persone studenti** e del personale di Ateneo, a qualsiasi titolo facente parte della comunità universitaria pisana, nonché favorire la realizzazione di un ambiente di studio e di lavoro inclusivo, sempre ispirato al valore fondante della pari dignità delle persone, eliminando situazioni di disagio per coloro che intendono modificare nome e identità nell'espressione della propria autodeterminazione di genere.

2. L'Università di Pisa, nel perseguimento degli obiettivi che ispirano il Comitato unico di garanzia "per le pari opportunità, la valorizzazione del benessere di chi lavora e contro le discriminazioni" (di seguito "CUG"), nel rispetto della normativa vigente, pone in essere le misure di protezione per le persone che abbiano la necessità di utilizzare, all'interno dell'Ateneo, un nome diverso rispetto a quello anagrafico, mediante l'attivazione di una carriera alias.

3. L'Università di Pisa, in accordo e attraverso le azioni positive del CUG, con l'obiettivo di garantire il pieno e positivo accesso al dispositivo della carriera alias, si impegna a fornire adeguata formazione in merito ai temi che riguardano l'identità di genere a tutto il personale coinvolto nelle procedure relative alla carriera alias e al personale in contatto diretto con **le persone** richiedenti.

Articolo 2 - Destinatari/ie

1. L'attivazione di una carriera alias può essere richiesta da tutti/e i/le componenti della comunità universitaria, docenti, studenti, personale tecnico-amministrativo, dirigenti, componenti esterni/e degli organi collegiali e quanti/e a vario titolo operano, anche occasionalmente e temporaneamente, nelle strutture dell'Ateneo.

2. Il/La delegato/a alle attività "Gender Studies and Equal Opportunities" può autorizzare il riconoscimento temporaneo di una identità alias in favore di ospiti dell'Ateneo che, in occasione di attività seminariali, convegni, cicli di lezioni o altre iniziative, intendano utilizzare su locandine, badge, ecc. un nome alias scelto in sostituzione dell'identità anagrafica ufficiale.

Articolo 3 - Procedura per l'attivazione carriera alias

1. L'attivazione della carriera alias avviene mediante presentazione al/la delegato/a alle attività "Gender Studies and Equal Opportunities" dell'istanza redatta secondo il modello allegato (All.to 1), corredata da un documento di identità.

2. Il/La delegato/a alle attività "Gender Studies and Equal Opportunities" assume le funzioni di tutor accademico per la carriera alias, supporta la persona richiedente nella procedura di richiesta e nella successiva gestione della carriera alias.

3. Il/La delegato/a, valutata l'istanza, acquisito il parere del/la Responsabile della Protezione Dati (di seguito "RPD") dell'Ateneo (RPDcarrieraalias@unipi.it), può autorizzare l'attivazione della carriera alias.

4. L'attivazione della carriera è subordinata alla sottoscrizione, da parte del/la Rettore/ rice e del/la Richiedente, di un Accordo di riservatezza, redatto secondo i modelli allegati (All.ti 2, 3 e 4).

Articolo 2 - Persone destinatarie

1. L'attivazione di una carriera alias può essere richiesta da **tutte le persone che compongono la comunità universitaria, ovvero persone docenti, persone studenti, personale tecnico-amministrativo, persone dirigenti, figure componenti esterne degli organi collegiali e da chiunque a vario titolo operi, anche occasionalmente e temporaneamente, nelle strutture dell'Ateneo.**

2. **La persona delegata** alle attività "Gender Studies and Equal Opportunities" può autorizzare il riconoscimento temporaneo di una identità alias in favore di persone ospiti dell'Ateneo che, in occasione di attività seminariali, convegni, cicli di lezioni o altre iniziative, intendano utilizzare su locandine, badge, ecc. un nome alias scelto in sostituzione dell'identità anagrafica ufficiale.

Articolo 3 - Procedura per l'attivazione carriera alias

1. L'attivazione della carriera alias avviene mediante presentazione **alla persona delegata** alle attività "Gender Studies and Equal Opportunities" dell'istanza redatta secondo il modello allegato (All.to 1), corredata da un documento di identità.

2. **La persona delegata** alle attività "Gender Studies and Equal Opportunities" assume le funzioni di **tutor d'ateneo** per la carriera alias, supporta la persona richiedente nella procedura di richiesta e nella successiva gestione della carriera alias.

3. **La persona delegata**, valutata l'istanza, acquisito il parere della **persona Responsabile della Protezione Dati** (di seguito "**persona RPD**") dell'Ateneo (RPDcarrieraalias@unipi.it), può autorizzare l'attivazione della carriera alias.

4. L'attivazione della carriera è subordinata alla sottoscrizione, **da parte del Rettore/ della Rettrice e della persona richiedente**, di un Accordo di riservatezza, redatto secondo i modelli allegati (All.ti 2, 3 e 4).

5. All'Accordo di riservatezza fa seguito un provvedimento del/la RPD contenente le istruzioni operative per gli uffici competenti.

Articolo 4 - Carriera alias

1. Al/la richiedente è assegnata un'identità provvisoria, transitoria e non consolidabile, al fine del rilascio di nuovi documenti di riconoscimento (libretto universitario - badge), di un nuovo account di posta elettronica e di targhette identificative, ove necessarie.
2. La carriera alias è inscindibilmente associata a quella riferita all'identità anagrafica legalmente riconosciuta e resta attiva per tutta la durata di quest'ultima, salve eventuali richieste dell'interessato/a o ipotesi di violazione del presente Regolamento.
3. I documenti di riconoscimento provvisori connessi alla carriera alias possono essere utilizzati esclusivamente all'interno dell'Università.
4. La carriera alias viene disattivata con il passaggio in giudicato della sentenza che, ai sensi della legge 14 aprile 1982 n. 164, attribuisca al/la richiedente un sesso diverso da quello enunciato nell'atto di nascita. L'anagrafica della carriera del/la richiedente viene quindi aggiornata sulla base del dispositivo della sentenza medesima. La carriera alias, in ogni momento, può essere disattivata a semplice richiesta dell'interessato/a.

5. All'Accordo di riservatezza fa seguito un provvedimento della **persona RPD** contenente le istruzioni operative per gli uffici competenti.

Articolo 4 - Carriera alias

1. Alla **persona richiedente** è assegnata un'identità provvisoria, transitoria e non consolidabile, al fine del rilascio di nuovi documenti di riconoscimento (libretto universitario - badge), di un nuovo account di posta elettronica e di targhette identificative, ove necessarie.
2. La carriera alias è inscindibilmente associata a quella riferita all'identità anagrafica legalmente riconosciuta e resta attiva per tutta la durata di quest'ultima, salve eventuali richieste **della persona interessata** o ipotesi di violazione del presente Regolamento.
3. I documenti di riconoscimento provvisori connessi alla carriera alias possono essere utilizzati esclusivamente all'interno dell'Università.
4. **La carriera alias viene disattivata con il passaggio in giudicato della sentenza che rettifica il genere anagrafico e i documenti d'identità della persona richiedente**²⁶. L'anagrafica della carriera della **persona richiedente** viene quindi aggiornata sulla base del dispositivo della sentenza medesima. La carriera alias, in ogni momento, può essere disattivata a semplice richiesta della **persona interessata**.

²⁶ Come specificato precedentemente, la sentenza di rettifica dei documenti può essere svincolata dalla riattribuzione chirurgica del sesso. Dato che all'università interessa la corrispondenza con il documento d'identità della persona richiedente per riaccorpere le due carriere (quella *alias* e quella abbinata al nome presente sui documenti all'atto dell'iscrizione), si può omettere il passaggio sulla riattribuzione del sesso.

Articolo 5 - Rilascio di certificazioni

1. Le attestazioni o certificazioni per uso esterno riguardanti la carriera alias, prodotte dall'Università, fanno esclusivo riferimento all'identità legalmente riconosciuta.
2. Il/La richiedente può rilasciare dichiarazioni sostitutive di certificazione o di atto notorio, relativamente a stati, fatti e qualità personali inerenti la carriera universitaria, da utilizzare all'esterno dell'Università esclusivamente con riferimento alla propria identità legalmente riconosciuta.

Articolo 6 - Obblighi del/la richiedente

1. Il/La richiedente si impegna a segnalare preventivamente e tempestivamente al/la RPD l'intenzione di compiere atti all'interno dell'Università con rilevanza esterna (a titolo esemplificativo, partecipazione a tirocini, adesione a progetti di mobilità internazionale, richiesta di borse di studio), impegnandosi a verificare e a concordare con il/la RPD stesso/a le modalità e le procedure di utilizzo della propria identità elettiva.
2. Il/La richiedente si impegna a comunicare all'Ateneo ogni mutamento in grado di influire sui contenuti e sulla validità dell'Accordo di riservatezza. In particolare, si impegna a comunicare tempestivamente il provvedimento di rettifica di attribuzione di sesso.

Articolo 7 - Violazione del Regolamento

1. In caso di violazione del presente Regolamento e dell'Accordo di riservatezza, la carriera alias può essere sospesa in via cautelare, all'esito di un'istruttoria preliminare, con provvedimento del/la Rettore/rice.
2. Qualora risulti accertata l'effettiva violazione, la carriera alias è disattivata con provvedimento del/la Rettore/rice, fatta salva l'irrogazione di eventuali successive sanzioni disciplinari e/o etiche.

Articolo 5 - Rilascio di certificazioni

1. Le attestazioni o certificazioni per uso esterno riguardanti la carriera alias, prodotte dall'Università, fanno esclusivo riferimento all'identità legalmente riconosciuta.
2. La **persona richiedente** può rilasciare dichiarazioni sostitutive di certificazione o di atto notorio, relativamente a stati, fatti e qualità personali inerenti la carriera universitaria, da utilizzare all'esterno dell'Università esclusivamente con riferimento alla propria identità legalmente riconosciuta.

Articolo 6 - Obblighi della persona richiedente

1. La **persona richiedente** si impegna a segnalare preventivamente e tempestivamente alla **persona RPD** l'intenzione di compiere atti all'interno dell'Università con rilevanza esterna (a titolo esemplificativo, partecipazione a tirocini, adesione a progetti di mobilità internazionale, richiesta di borse di studio), impegnandosi a verificare e a concordare con la **persona RPD stessa** le modalità e le procedure di utilizzo della propria identità elettiva.
2. La **persona richiedente** si impegna a comunicare all'Ateneo ogni mutamento in grado di influire sui contenuti e sulla validità dell'Accordo di riservatezza. In particolare, si impegna a comunicare tempestivamente **il provvedimento di rettifica dei documenti d'identità**.

Articolo 7 - Violazione del Regolamento

1. In caso di violazione del presente Regolamento e dell'Accordo di riservatezza, la carriera alias può essere sospesa in via cautelare, all'esito di un'istruttoria preliminare, con provvedimento **del Rettore/della Rettrice**.
2. Qualora risulti accertata l'effettiva violazione, la carriera alias è disattivata con provvedimento **del Rettore/della Rettrice**, fatta salva l'irrogazione di eventuali successive sanzioni disciplinari e/o etiche.

Articolo 8 - Accordo di riservatezza

1. L'Accordo di riservatezza ha efficacia dalla data di sottoscrizione e cessa immediatamente al momento del venir meno dei presupposti che lo hanno determinato.
2. Nel caso in cui lo/la studente richiedente consegua il titolo di studio senza l'intervento di un provvedimento di cambiamento di genere o di identità, gli atti di carriera si intendono riferiti ai dati anagrafici effettivi del/la richiedente e conseguentemente sono rilasciate le certificazioni e i documenti con i dati personali della carriera anagrafica.
3. I provvedimenti relativi alla carriera del personale dipendente dell'Ateneo sono emessi con riferimento all'identità alias solo se destinati a uso interno, mentre la documentazione destinata a uso esterno è unicamente riferita all'identità legalmente riconosciuta.

Articolo 9 - Testimoni e collaboratori/rici di giustizia

1. L'attivazione di una carriera alias può essere altresì richiesta da:
 - testimoni e collaboratori/rici di giustizia;
 - chiunque sia sottoposto dalle competenti autorità a una misura di protezione che prevede il cambio di identità per motivi di sicurezza;
 - autorità competenti in materia;
 - chiunque denunci delitti contro la pubblica amministrazione o delitti compiuti mediante abuso di autorità o dei poteri inerenti la pubblica funzione o il pubblico servizio.
2. Nelle ipotesi di cui al comma 1, si applicano, in quanto compatibili, le

Articolo 8 - Accordo di riservatezza

1. L'Accordo di riservatezza ha efficacia dalla data di sottoscrizione e cessa immediatamente al momento del venir meno dei presupposti che lo hanno determinato.
2. Nel caso in cui **la persona richiedente** consegua il titolo di studio senza l'intervento di un provvedimento di cambiamento di genere o di identità, gli atti di carriera si intendono riferiti ai dati anagrafici effettivi **della persona richiedente** e conseguentemente sono rilasciate le certificazioni e i documenti con i dati personali della carriera anagrafica.
3. I provvedimenti relativi alla carriera del personale dipendente dell'Ateneo sono emessi con riferimento all'identità alias solo se destinati a uso interno, mentre la documentazione destinata a uso esterno è unicamente riferita all'identità legalmente riconosciuta.

Articolo 9 - Testimoni e collaboratori/ collaboratrici di giustizia²⁷

1. L'attivazione di una carriera alias può essere altresì richiesta da:
 - testimoni e collaboratori/ collaboratrici di giustizia;
 - **qualsunque persona sia sottoposta** dalle competenti autorità a una misura di protezione che prevede il cambio di identità per motivi di sicurezza;
 - autorità competenti in materia;
 - chiunque denunci delitti contro la pubblica amministrazione o delitti compiuti mediante abuso di autorità o dei poteri inerenti la pubblica funzione o il pubblico servizio.
2. Nelle ipotesi di cui al comma 1, si applicano, in quanto compatibili, le

²⁷ In questo caso è opportuno lasciare la forma binaria poiché *collaboratore/collaboratrice di giustizia* è un'espressione polirematica che indica un preciso status regolato da apposita normativa.

disposizioni del presente Regolamento, con le seguenti precisazioni:

- l'istanza è presentata al Rettore/ricе;
- il procedimento si svolge in piena collaborazione con l'autorità competente;
- le caratteristiche dell'identità sono determinate dall'autorità competente in materia o in accordo con essa;
- i documenti di riconoscimento provvisori connessi alla carriera alias possono essere utilizzati esclusivamente all'interno dell'Università, con l'eccezione di quanto disposto dalle autorità competenti;
- in caso di violazione del presente Regolamento e dell'Accordo di riservatezza, ogni provvedimento di cui all'art. 7 è adottato di concerto con l'autorità competente.

Articolo 10 - Obblighi di segretezza

1. L'Università tratterà i dati indicati nell'accordo di riservatezza secondo quanto disposto dalla normativa vigente.
2. Il personale dell'Università che interviene nel procedimento relativo alla carriera alias è tenuto alla segretezza delle informazioni acquisite.

Articolo 11 - Entrata in vigore e pubblicità

1. Il Regolamento è emanato con decreto del/la Rettore/ricе ed entra in vigore 15 giorni dopo la sua pubblicazione sul sito web di Ateneo nell'Albo ufficiale Informatico.
2. È altresì reso pubblico sul sito web di Ateneo e sulla pagina ufficiale del CUG.

disposizioni del presente Regolamento, con le seguenti precisazioni:

- l'istanza è presentata **al Rettore/alla Rettrice**;
- il procedimento si svolge in piena collaborazione con l'autorità competente;
- le caratteristiche dell'identità sono determinate dall'autorità competente in materia o in accordo con essa;
- i documenti di riconoscimento provvisori connessi alla carriera alias possono essere utilizzati esclusivamente all'interno dell'Università, con l'eccezione di quanto disposto dalle autorità competenti;
- in caso di violazione del presente Regolamento e dell'Accordo di riservatezza, ogni provvedimento di cui all'art. 7 è adottato di concerto con l'autorità competente.

Articolo 10 - Obblighi di segretezza

1. L'Università tratterà i dati indicati nell'accordo di riservatezza secondo quanto disposto dalla normativa vigente.
2. Il personale dell'Università che interviene nel procedimento relativo alla carriera alias è tenuto alla segretezza delle informazioni acquisite.

Articolo 11 - Entrata in vigore e pubblicità

1. Il Regolamento è emanato con decreto **del Rettore/della Rettrice** ed entra in vigore 15 giorni dopo la sua pubblicazione sul sito web di Ateneo nell'Albo ufficiale Informatico.
2. È altresì reso pubblico sul sito web di Ateneo e sulla pagina ufficiale del CUG.

Bibliografia

- Bombi, Raffaella, 2013, *Manuale di comunicazione istituzionale*, Roma, Il Calamo.
- Bombi, Raffaella, 2015, *Quale comunicazione tra stato e cittadino? Per un nuovo manuale di comunicazione istituzionale e internazionale*, Roma, Il Calamo.
- Brunato, Dominique, 2014, “Complessità necessaria o stereotipi del “burocratese”? Un’indagine sulla leggibilità del linguaggio amministrativo da una prospettiva linguistico-computazionale”, in Ruffino G. & Castiglione M. (a cura di), *La lingua variabile nei testi letterari, artistici e funzionali contemporanei: analisi interpretazione, traduzione*, Atti del XII Congresso della Società internazionale di Linguistica e Filologia Italiana – SILFI (Palermo, 22-24 settembre 2011), Centro studi filologici e linguistici siciliani, Firenze, Cesati (<http://www.italianlp.it/wp-content/uploads/2014/10/articolo_silfi_2014_versione_definitiva.pdf>).
- Brunato, Dominique & Venturi, Giulia, 2014, “Le tecnologie linguistico-computazionali nella misura della leggibilità di testi giuridici”, in Tiscornia D., Romano F. & Sagri M. T. (a cura di), *Diritto, linguaggio e tecnologie dell’informazione*, fascicolo monografico di *Informatica e diritto* 1, 111-142.
- Brunato, Dominique & Venturi, Giulia, 2016, “Le tecnologie del linguaggio per la leggibilità della comunicazione istituzionale”, in Panizza S. (a cura di), *Profili attuali di qualità degli atti normativi e amministrativi*, Pisa, Pisa University Press, 123-161.
- Cassese, Sabino, 1983, “Il linguaggio della burocrazia”, in *Il linguaggio della divulgazione, Atti del secondo convegno nazionale promosso da Reader’s Digest (Roma, Accademia dei Lincei, 14-15 aprile 1983)*, Milano, Selezione Reader’s Digest, 42-48.
- Comandini, Gloria, 2021, “Salve a tuttə, tutt*, tuttu, tuttx e tutt@: l’uso delle strategie di neutralizzazione nella comunità queer online. Indagine su un corpus di italiano scritto informale sul web”, *Testo e senso* 23, 43-64.
- Cortelazzo, Michele, con la collaborazione di Pellegrino, Federica & Viale, Matteo, 1999, *Semplificazione del linguaggio amministrativo. Esempi di scrittura per le comunicazioni ai cittadini*, Padova, Comune di Padova.
- Cortelazzo, Michele, 2014a, “L’italiano nella scrittura amministrativa”, in Lubello S. (a cura di), *Lezioni d’italiano. Riflessioni sulla lingua del nuovo millennio*, Bologna, Il Mulino, 85-104.
- Cortelazzo, Michele, 2014b, *Modelli di comunicazione standard nelle università. I. Personale tecnico e amministrativo*, Padova, CLEUP.
- Cortelazzo, Michele, 2015, *La semplificazione dei testi amministrativi. Le buone pratiche*, in Bombi R., 2015, *Quale comunicazione tra stato e cittadino? Per un nuovo manuale di comunicazione istituzionale e internazionale*, Roma, Il Calamo, 93-110.
- Cortelazzo, Michele, 2021, *Il linguaggio amministrativo. Principi e pratiche di modernizzazione*, Roma, Carocci.

- Cortelazzo, Michele & Pellegrino, Federica, 2002, "Trenta regole per scrivere testi amministrativi chiari", in *Guida agli enti locali*, 25 maggio 2002 (20), XXV-XXXV.
- Cortelazzo, Michele & Pellegrino, Federica, 2003, *Guida alla scrittura istituzionale*, Roma-Bari, Laterza.
- Dell'Orletta, Felice, Montemagni, Simonetta & Venturi, Giulia, 2011, "READ-IT: Assessing Readability of Italian Texts with a View to Text Simplification", in *SLPAT '11 - Proceedings of the Second Workshop on Speech and Language Processing for Assistive Technologies (Edinburgh, July 30th 2011)*, Stroudsburg, PA, USA, Association for Computational Linguistic, 73-83.
- Dell'Orletta, Felice, Montemagni, Simonetta & Venturi, Giulia, 2014, "Assessing Document and Sentence Readability in Less Resourced Languages and across Textual Genres", in *Recent Advances in Automatic Readability Assessment and Text Simplification. Special issue of International Journal of Applied Linguistics*, 165 (2), John Benjamins Publishing Company, 163-193.
- Fioritto, Alfredo, 1997, *Manuale di stile. Strumenti per semplificare il linguaggio delle amministrazioni pubbliche*, Bologna, Il Mulino.
- Fortis, Daniele, 2005, "Il linguaggio amministrativo italiano", *Revista de Liengua i dret* 43, 47-116.
- Franceschini, Fabrizio & Gigli, Sara (a cura di), con la collaborazione di Annecker, Marco, Bellantoni, Manuela, Caldaroli, Paola & Femia, Rita, 2003, *Manuale di scrittura amministrativa*, Università di Pisa, Dipartimento di Studi italianistici, con la collaborazione dell'Agenzia delle Entrate (<<http://www.segretariientilocali.it/nuovo/File2003/ManualeScritturaAmministrativa.pdf>>).
- Gotti, Maurizio, 2005, *Investigating Specialized Discourse*, Bern, Peter Lang.
- GRADIT = De Mauro, Tullio (a cura di), 1999, *Grande dizionario italiano dell'uso*, Torino, UTET.
- Gualdo, Riccardo & Telve, Stefano, 2011, *Linguaggi specialistici dell'italiano*, Roma, Carocci.
- Kruppa, Alexandra, Fenn, Julius & Ferstl, Evelyn C., 2021, *Does the asterisk in gender-fair word forms in German impede readability? Evidence from a lexical decision task*, report presented within the conference *Architectures and Mechanisms for Language Processing - AMLaP (Paris, September 2nd-4th 2021)*. Complete video-recording of the report available here: <<https://www.youtube.com/watch?v=BCJoS1taGXU>>.
- Lubello, Sergio, 2014, *Il linguaggio burocratico*, Roma, Carocci.
- Lubello, Sergio, 2017, *La lingua del diritto e dell'amministrazione*, Bologna, Il Mulino.
- Lucisano, Pietro & Piemontese, Maria Emanuela, 1988, "Gulpease. Una formula per la predizione della difficoltà dei testi in lingua italiana", *Scuola e città* 1988 (3), 57-68.

- Montemagni, Simonetta, 2013, “Tecnologie linguistico-computazionali e monitoraggio della lingua italiana”, *Studi Italiani di Linguistica Teorica e Applicata* XLII (1), 145-172.
- Mortara Garavelli, Bice, 1997, *Manuale di retorica*, Milano, Bompiani, nuova edizione ampliata (I edizione: 1988).
- Robustelli, Cecilia, 2012, *Linee guida per l'uso del genere nel linguaggio amministrativo*, Firenze, Accademia della Crusca e Comune di Firenze.
- Robustelli, Cecilia, 2015, “Tra politica e linguistica. Genere grammaticale e ruoli istituzionali”, in Accademia della Crusca (a cura di), *Atti del convegno “L'italiano istituzionale nella UE e in Italia” (Roma, 20 maggio 2014)*, 45-56.
- Robustelli, Cecilia, 2016, “Linguaggio discriminatorio e testi istituzionali”, in Panizza S. (a cura di), *Profili attuali di qualità degli atti normativi e amministrativi*, Pisa, Pisa University Press, 99-122.
- Robustelli, Cecilia, 2021, “Linee guida per l'uso del genere nel linguaggio amministrativo con appendice di testi discussi durante un corso di formazione presso il Comune di Parma e proposta di riscrittura”, in Dalcò, Fabrizia (a cura di), *Genere e comunicazione istituzionale*, Progetto formativo del Comune di Parma, Cosenza, Arti Grafiche Cardamone srl, 1-55 (<https://iris.unimore.it/retrieve/e31e124f-efdb-987f-e053-3705fe0a095a/Genere%20e%20comunicazione%20istituzionale%202021.pdf>).
- Thornton, Anna M., 2022, “Genere e igiene verbale: l'uso di forme con ə in italiano”, *Annali del Dipartimento di Studi Letterari, Linguistici e Comparati (AION) - Sezione Linguistica* 11, 11-54.
- Vellutino, Daniela, 2018, *L'italiano istituzionale per la comunicazione pubblica*, Bologna, Il Mulino.
- Viale, Matteo, 2008, *Studi e ricerche sul linguaggio amministrativo*, Padova, CLEUP.
- Voghera, Maria (Miriam), 2005, “La misura delle categorie sintattiche”, in Chiari I. & De Mauro T. (a cura di), *Parole e numeri. Analisi quantitative dei fatti di lingua*, Roma, Aracne, 125-138.

NICOLA SANTONI*

In search for the motivation of markedness through metaphorical processes. Chains and matrices with a reasonable degree of universality

Abstract

After Greenberg's studies on implicational universals, dealing with universals of language is dealing with markedness. Interestingly enough, (un-)marked linguistic forms can show their motivation if examined at the conceptual level. On the Cognitive Linguistics ground, Metaphor Theory and Image Schemas are here employed to search for that motivation (if any), and to verify if semantics could be behind forms traditionally considered a mere fact of grammar.

Keywords: Metaphor, Semantics, Universals, Cognitive Linguistics, Prosodic Factors

A partire dagli studi di Greenberg sugli universali implicazionali, affrontare il tema degli universali linguistici significa avere a che fare anche con la marcatezza. È interessante notare che forme linguistiche (non) marcate possono rivelare la propria motivazione, se vengono esaminate al livello concettuale. Con le prospettive offerte dalla linguistica cognitiva, la teoria della metafora e gli schemi di immagini sono qui impiegati per scoprire quella motivazione, se presente, e per verificare se ragioni semantiche siano alla base di forme tradizionalmente considerate puri fatti sintattico-grammaticali.

Parole chiave: Metafora, Semantica, Universali linguistici, Linguistica cognitiva, Fattori prosodici

1. Universals of language and cognitive mechanisms

The articulated set of theories we commonly refer to as Cognitive Linguistics has undoubtedly proven itself an important tool to deal with general and specific semantic mechanisms of language. It often demonstrated that semantic-psychological operations show various degrees of universality, and they are rooted in experiential bases, shared by the vast majority of human beings, that give motivation to linguistic form. Among the most productive

* Nicola Santoni, Università di Macerata, n.santoni1@unimc.it.

theories, Metaphor Theory surely stands out in the complex topic dealing with universals of language. The main issue here is if conceptual metaphors (as we know them since the works, at least, of G. Lakoff and M. Johnson) can be assigned a role in the determination of marked and unmarked forms, in spite of a traditional and modular view (nowadays mostly criticized in various aspects by many linguists) which tends to separate levels of analysis of language, and to treat them as independent components¹.

Markedness has played a central role since the very beginning of the research on universals; after Greenberg, it has become a hard task to refer to them ignoring this key-feature of linguistic forms. Joseph Greenberg, as a well-known fact, took inspiration from the phonological models of Trubetsky and Jakobson, and adopted the implicational logical form to discuss the behaviour of universals across language types (Greenberg 1963; 1966). We must bear in mind that the general rule states that “Given x in a particular language, we always find y. When nothing further is said, it is understood that the converse, namely ‘given y, we always find x’, does not hold”². In addition, it has to be underlined that the markedness parameter is explicitly conceived by Greenberg as gradient, relational, non-strictly dichotomic³, and it acquires great importance in the research programme of Greenberg (1966)⁴. In third place, we know that implicational universals work usually at the same linguistic level and in a one-directional way on the basis of markedness. To give an example,

It will be found that in generalizing statements regarding sound sequences it is usually the unmarked feature which figures in the *implicatum* of conditional statements. Thus in the statement that the existence of clusters containing at least one glottalized member implies the existence of clusters containing

*

¹ See Santoni 2020 for an earlier inquire on this topic.

² Greenberg 1963: 73.

³ For instance, “in regard to the category of number in the noun [...] there is much evidence for a hierarchy singular, plural, dual from the most unmarked to the most marked” (Greenberg 1963: 31). Likewise in this article, each semantic-cognitive domain is not conceived as absolutely marked or unmarked (regardless the relations with other contexts and domains), but always relatively.

⁴ “The topic of universals is here approached through the consideration with a single, but [...] rich and complex set of notions, those pertaining to marked and unmarked categories” (Greenberg 1966: 10).

exclusively non-glottalized members, it is the unmarked feature, non-glottalized, which is the implied one⁵.

Despite this, there is much more in Greenberg than the simple intuition that relations between marked forms could cross the boundaries of hypothetical modular linguistic levels. Interestingly, “sometimes the marked category in phonology is the expression of a marked category in grammar. Thus certain Amerind languages use the marked feature of glottalization to express the marked grammatical category of the diminutive” (Greenberg 1966: 69).

In this framework, it seems reasonable to ask ourselves if there could be cognitive and semantic motivation behind markedness relations, and maybe if (un-)marked domains could imply motivated (un-)marked correspondent forms, which are more universal, the more universal is the mapping process motivating them. In a cognitive perspective, it has to be recalled that “Grammatical universals are universals concerning the pairing of form and content; they are not universals of form alone (whatever that could mean). [...] Linguistic universals include conceptual universals (e.g., primitive spatial relations, universal conceptual metaphors), universals of cognitive function, and universals of iconicity”⁶.

2 Metaphorical chaining. Type [S_{unm} > T > T1_{unm}]

If we look at the syntactic level, we remember that “Universal 1” clearly defines the universal basic linear order (BO) as the one in which the role Subject precedes the role Direct Object in the unmarked declarative propositions. It has been suggested by Langacker (1991) that this fact is due to the image schema of the so-called “Billiard Ball Model” (BBM), on which the “Canonical Event Model” (CEM) is shaped and metaphorized, on the basis of the “Iconicity Principle” of grammar⁷. This would be the starting point experientially motivating the form of BO at a prelinguistic level. In other words, the physical and perceptual experience of force dynamics suggests us that energy transfers (in the most general sense) typically go from an Agent at t_0 to a Patient at t , and it represents

⁵ Greenberg 1966: 22.

⁶ Lakoff & Johnson 1999: 506.

⁷ For details and specific applications, see Serra Borneto 2003 as a good example.

the unmarked default for cognition. It plausibly follows that the order of appearance of semantic roles in the linguistic form constitutes a metaphorical mapping of the prototypical transitive event, which is in turn mapped onto the already mentioned BBM. Some clues about unmarkedness of this image schema come from the fact that it can propagate itself even in syntactic structures which involve other roles or indirect causality relations⁸. That is tantamount to say that what appears in “metaphorical neutralization” on the surface is the default schema⁹, as we should expect from a neutralization. We can try to summarize this situation with a simple linear scheme, representing a metaphorical mapping that goes from a source domain to the target domains:

- 1)
[BBM (S_{unmm}) > CEM (T) > Basic order (T_{1unmm})]

The unmarked Source domain is the basis of the mapping for the transitive event Target, on which the formal result Target₁ is mapped, namely the syntax of the unmarked transitive active construction. It can be easily recognized that the mapping structure is a chain of metaphors (chaining process), and that unmarkedness pertaining to Source domain is linearly transferred until it reaches the form of Target₁ domain.

2.1 Spatial neutralization and chaining

Similarly, spatial constructions indicating emotional or physical states and conditions¹⁰ provide another example of metaphorical mapping, which motivates the linguistic form resulting in metaphorical neutralization. If we look at Italian, there are two possible and alternative constructions for a number of conditions, like *Essere in ansia/Essere ansioso*, *Essere in depressione/*

⁸ It. *Marco ha colpito Paolo/Marco ha visto Paolo/Marco ha provocato la nostra reazione* have identical (morpho-)syntax at the linear and structural level, although in the second case *Marco* has an EXPeriential role, and in the third both causality relation and energy transfer are indirect. Please notice that this simple example taken from Italian could easily hold for most nominative-accusative languages.

⁹ Moreover, marked passive constructions are often the mirror-images of transitive unmarked active constructions (Taylor 2003 [1995]: 333-334), if we look at the linear level at least for the mentioned language typology (n. 8).

¹⁰ See Lakoff & Johnson 1980: 30-32 for this metaphor and some examples in English.

Essere depresso, Essere in tempesta/Essere tempestoso, which of course can't be considered couples of perfect synonyms. There are also constructions showing no alternative but the spatial-prepositional one, almost totally excluding the predicative synonym-like form with the adjective: *Essere in forma/Ø, Essere in gara/Ø, Essere in crisi/Ø, Essere in pericolo/Ø*, and many others¹¹. No matter here what causes the blocking (or when exactly it is caused) in these second cases, all examples regardless share the spatial form [IN__] in neutralization for emotional domains or states metaphorized in space, with slight (yet not irrelevant) semantic consequences if alternatives are given. If we look at Mandarin Chinese, we find very close examples of this phenomenon, which seems to have a high degree of universality. There are cases like

Tā zhèng chǔyú bēishāng zhōng / Tā hěn bēishāng¹²
 She now be sadness in / She very sad

and also metaphorical neutralizations lacking the predicative alternative, as in

Guójiā chǔyú wēijī (zhì)zhōng / Ø
 State be crisis in / Ø

¹¹ Sometimes, where lexicon and morphology allow to attempt a substitution of spatial form with an adjective in these cases, the overlap between hypothetical alternatives is impossible. Thus, for example, *Essere in pericolo* (*To be in danger*) means the “passive” condition of a [+ Animated] subject, while *Essere pericoloso* (*To be dangerous*), if referred to [+Animated], means quite the “active” opposite. Almost the same happens with *Essere in crisi* (*To be in crisis*) and *Essere critico* (*To be critical, decisive, crucial*), even referred to [+ Animated]. When it comes to [- Animated], things become more complex to discuss in detail, but it should be pointed out that some degree of synonymy can't be overruled. Take the special case of *life: La (sua) vita è in pericolo/La (sua) vita è pericolosa* (*His life is in danger/His life is dangerous*) displays almost the same semantic proximity of the alternatives as in the examples of the kind *Essere in depressione/Essere depresso* (*To be in a depression/To be depressed*), in which the difference of construction (metaphorical spatial vs predicative) marks sometimes the opposition “temporary condition vs permanent state”.

¹² In Mandarin a positive adjective can hardly stand alone in predicative structures. Thus, the sentence needs what we consider the comparative structure with “*hěn* + adjective” to be well-formed. Otherwise, if spatial metaphorization with “*zhōng*” is involved, there is no need for further elements, since “*bēishāng*” (and adjectives in general) can work equally as a noun, because nouns and adjectives share the same morphology. One could be tempted to discuss if “*hěn*” works as a *copula* in these structures, but this would bring us too far from our discourse.

Spatial construction [___ZHŌNG] appears in neutralization. Here are some more examples from Mandarin:

X [Somebody] chǔyú juéwàng (zhī)zhōng / X hěn juéwàng
 (X be despair in / X very desperate)
 Macerata chǔyú píngjìng (zhī)zhōng / Macerata hěn píngjìng
 (Macerata be quiet in / Macerata very quiet)
 X chǔyú hàipà (/kǒngjù*)(zhī) zhōng / X hěn hàipà
 (X be fear in / X very frightened)
 X chǔyú wēixiǎn zhōng / Ø
 (X be danger in / Ø)
 Guójiā chǔyú jùbiàn zhōng / Ø
 (State be transformation in / Ø)

The metaphorical chaining structure seems to be the same as 1), and in this case it involves Physical Space and Container domains to map State (or condition):

2)
 [Container (S_{unm}) > Space (T) > State/Condition (T_{1 unm})]

Metaphorical neutralization of states in containers rests upon cognitive metaphors operating with a high degree of universality for Cognitive Linguistics. In this case, the ontological metaphors mainly involved are SPACES ARE CONTAINERS, and STATES/CONDITIONS ARE SPACES, with all the boundaries and spatial characteristics conceptualized on containers¹³. We observe that the unmarkedness of the well-known *Container Schema* (for it is a fundamental and default schema for our cognitive system as human beings to conceptualize many everyday experiences) is transferred to the end of the chaining process, in the unmarked T₁ construction, which neutralizes actual spaces and metaphorical conditions in the spatial (and, in these cases, pre- or postpositional locative) form across languages.

¹³ We know from conceptualized experience that we can enter or exit spaces/containers, or we can be trapped inside them. We know that they have limits, and their boundaries divide an internal from an external part; also, we are aware that each container/space has a capacity: if objects inside are too many or too big, they are tightly contained, and freedom of movement is highly reduced for them.

3. Metaphorical matrices. Type $[S_{\text{unm}} > (T + T_1) > T_{2M}]$

At the suprasegmental phonological level, Greenberg (1966: 70) noticed that “Another example of phonological-grammatical connection is the widespread use of the marked category of final rising pitch for the expression of interrogation”. This relation between rising pitch and interrogation of the *yes/no* kind leads us to have a deeper look into the three main prosodic factors, namely pitch, prominence, and length. In the perspective here adopted, it seems that the three factors play a complex role if they take part to the semantic motivation orienting markedness of linguistic forms. If that’s the case, everything starts from unmarked cognitive extralinguistic domains as well. The metaphorical underlying structure in these circumstances is more complex than the chain, as it’s going to be cleared by what follows.

3.1 Pitch variation

The orientational metaphor here involved can be indicated by UNKNOWN IS UP¹⁴ (and KNOWN IS DOWN). To verify the cross-linguistic persistence of this cognitive image, we must go through lexicon and idiomatic expressions (which are rarely perceived as actual metaphors by speakers).

It. *È sotto gli occhi di tutti*¹⁵

(something known is “under everyone’s eyes”)

It. *Un discorso terra terra*

(a discourse so simple to understand that is “at the ground level”)

Mand. *Chénāi luò dìng*

(“Dust has already fallen to the ground”, the fact is concluded and the results are known)

Mand. *Jǐnguǎn māma jiěshì le suǒyǒu, wǒ réngrán juéde yúnliǔwùlǐ*¹⁶

(“Even if mom explained everything to me, I’m still between clouds and fog”)

¹⁴ See Lakoff & Johnson 2012 [1980]: 39–40; 172. Several English examples of this metaphor can be found in Lakoff & Johnson 1980. For this reason, I don’t largely discuss them in this article, because Metaphor Theory was born mainly in this language, and the founders’ influential works in Cognitive Linguistics have already explained them in full detail.

¹⁵ This example is also structured by the metaphor KNOWING IS SEEING, as the last Mandarin one referring to “clouds and fog”.

¹⁶ For specific implications of this and other metaphors in Chinese, see Yu 1998; 2009.

Let's consider now the cognitive spatial domain of HEIGHT/VERTICALITY as an unmarked basic domain, through the clues of unmarkedness via neutralization.

(It.) *Quanto sei alto/*basso?*

(It.) *Sono alto/*basso solo un metro e cinquanta*

(It.) *L'altezza/la *bassezza viene calcolata con formule matematiche*

(Eng.) *How tall/*short are you?*

(Eng.) *I'm only 4 ft tall/*short*

(Eng.) *Height/*lowness measurement*

(Eng.) *He's averagely-tall/*short*

(Eng.) *To have greater (or lower) height/*lowness*

(Mand.) *Nǐ duō gāo/*dī?* ("How tall are you?"; gāo = adj. "tall", dī = adj. "short")

(Mand.) *Jìsuàn gāo dù/∅* ("To measure height"; gāo dù = "height")

As Lakoff & Johnson (2012 [1980]) already observed, the motivation upon which the widespread¹⁷ rising intonation in marked *yes/no* questions rests should be ultimately connected to the above mentioned metaphor. Yet, they tend to exclude this possibility in Chinese and in tone-systems, since these languages make no use of final rising pitch, because they are constrained by pre-assigned tones (*contra*, see e.g. De Dominicis 2013: 337-338). But this could be only a part of the story. An experimental study on Mandarin speakers suggested that even in this language rising intonation could emerge in particular circumstances to mark the *yes/no* questions, even if they have the final particle "ma" (traditionally considered bearing the neutral tone) signaling the sentence type.

3.1.1 Case study: Rising intonation as a useful tool in Chinese?

In this experiment¹⁸, ten Mandarin mother-tongue speakers were asked to read the following questions, written in *pinyin*. At first, they were recorded reading these sentences at normal speed, then they were requested to read

¹⁷ As a matter of fact, rising pitch in *yes/no* questions is at disposal of a great number of different languages, but "non si tratta di un universale e non tutte le lingue usano il tono per segnalare l'interrogazione. È il caso, ad esempio, di alcune lingue del gruppo niger-congo (famiglie gur, kwa e kru) e del gruppo nilotico-sahariano (lingue centro-sudanesi e ciadiche del gruppo afro-asiatico) [...], o del dialetto di Bomarzo, una località dell'alta Tuscia, a nord di Viterbo" (De Dominicis 2010: 34).

¹⁸ Data recorded and analyzed between May and September 2019, and here published for the first time.

them again slowly to make a foreigner (average Chinese speaker) understand, and they were recorded again.

1-Nǐ lái zì GuǎngZhōu ma? (“Do you come from Guangzhou?”)

2-Duì ma? (“Yes?”)

3-Zhēn de ma? (“Really?”)

4-Nǐ kàn guò ma? (“Have you ever seen?”)

5-Nǐ chī guò ma? (“Did you eat?”)

C-Nǐ duō gāo? (“How tall are you?”)

“C” stands for “Control”, because it is the only non-yes/no question bearing actually tone 1 in the end (“gāo”), the tone of maximum height. Therefore, the comparison was made between the “ma” frequencies of the other questions, both at normal speed and hyperarticulation, and C frequencies. Using a recording and editing free software, frequency peaks were all calculated for each sentence-ending, and the average peak at normal speed (AVG_N) was obtained multiplying C by 5 (the number of the yes/no questions), then adding for each speaker (Sp.) the other 5 normal speed peaks, and finally dividing the result by 10 for each speaker. This was made to verify if final intonation shows rising pitch in these questions beyond the average, and also if it equals or exceeds the tone 1 frequencies in hyperarticulation for each participant. Of course, for each speaker was equally calculated the average frequency of tone 1 in C (C_{AVG}), adding normal speed frequency (N) to hyperarticulated mode (H), and dividing the result by 2. Before going to the results, it has to be reported that for female (F) voices tone 1 is around 250-290 Hz, while for males (M) the frequency is around 170-210 Hz.

3)

Questions	M	F	F	M	F	F	F	M	M	M
	Sp. 1 N / H (Hz)	Sp. 2 N / H (Hz)	Sp. 3 N / H (Hz)	Sp. 4 N / H (Hz)	Sp. 5 N / H (Hz)	Sp. 6 N / H (Hz)	Sp. 7 N / H (Hz)	Sp. 8 N / H (Hz)	Sp. 9 N / H (Hz)	Sp. 10 N / H (Hz)
1 (ma)	209/ 223	195 / 208	238/ 258	161 / 141	260/ 258	217/ 234	226 / 240	180 / 203	156/ 151	138/ 145
2 (ma)	225 / 199	160/ 188	206/ 201	139/ 114	245/ 254	205/ 215	220/ 224	173/ 191	158/ 157	142/ 120
3 (ma)	230 / 246	155/ 172	239/ 238	172 / 148	226/ 230	215/ 224	234 / 243	171/ 205	153/ 154	148/ 136

4 (<i>ma</i>)	181/ 171	152/ 159	191/ 183	113/ 113	224/ 223	198/ 223	186/ 180	153/ 167	131/ 152	110/ 123
5 (<i>ma</i>)	196/ 176	149/ 151	210/ 236	147/ 121	215/ 228	210/ 220	193/ 177	147/ 164	146/ 157	146/ 119
C (<i>gāo</i>)	225/ 218	172/ 194	258/ 265	169/ 150	263/ 260	257/ 292	227/ 219	169/ 191	170/ 191	166/ 154
C _{AVG} tone 1	<u>221,5</u>	<u>183</u>	<u>261,5</u>	<u>159,5</u>	<u>261,5</u>	<u>274,5</u>	<u>223</u>	<u>180</u>	<u>180,5</u>	<u>160</u>
AVG _N	216,6	167,1	237,4	157,7	248,5	233	219,4	166,9	159,4	151,4

A) The numerous cases highlighted in grey show the increment to a higher frequency for the hyperarticulated mode with some degree of emphasis by the speakers. Also the C question is affected by this general increment, a sign of a widespread use of final rising pitch in H-context.

B) Underlined and in bold are all the peaks which equal or exceed the C_{AVG} of tone 1. This happens in 18 cases (8 N and 10 H) limited to the first 3 questions. In questions 1 and 3 this could be quite predictable, since co-articulation of “*ma*” with a previous tone 1 can favour the phenomenon. But in 2, which displays a significant rise, “*dui*” can’t be responsible for this, and the reason has probably to be found in semantic features: 2 is really a prototypical *yes/no* question (“*Yes?*”), a fact that probably clearly emerges in intonation.

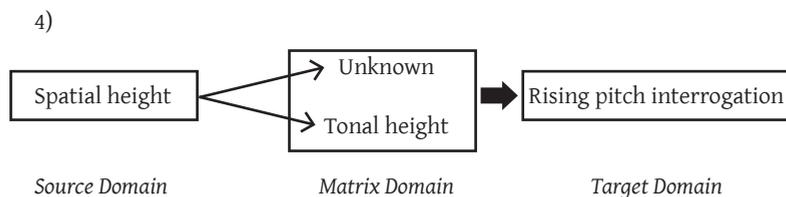
C) In 5 cases, just underlined, we have values which equal or exceed at least one of the C peaks, normal or hyperarticulated. Of course, values of C_{AVG} are all superior to one of the C peaks, and were underlined as a reference parameter. Also in these cases, regarding questions 1, 2, and 3 as well, the tone of “*ma*” tends to reach tone 1, but less evidently than in the previous 18 cases.

D) We have in total 23 cases in the first 3 questions that tend to tone 1, representing more than 38% in the group 1-3, and 23% relatively to 1-5. In 4 and 5, we have no trace of tone 1, but they follow the tendency of rising pitch in the passage from N to H mode.

It seems quite evident that final rising pitch and tone 1 is at disposal of Mandarin to mark (*yes/no*) questions with a number of occurrences far from chance. Rather, rising intonation is *also* one of the “tools” Mandarin can put in play in particular circumstances, like emphasis or general hyperarticulation.

3.1.2 Unknown, height, and interrogation

At this point, we need to hypothesize the existence of another ontological metaphor to fully explain the relation between unknown, height/verticality, and the linguistic form of rising pitch in *yes/no* questions. The basis of this metaphor is strictly anatomical and experiential, and scarcely testified by lexical clues. Nonetheless, it seems plausible to state that for our cognitive system TONAL HEIGHT IS SPATIAL HEIGHT¹⁹. If so, we come to a completely different mapping structure from 1) and 2). The presence of a matrix domain²⁰ (or “blended space”²¹) plays a key role for the possible motivation of linguistic form in this and the following cases.



From an unmarked S two mappings proceed towards the target domains we can refer to as T and T_1 . The new complex domain (the matrix domain) is the core of the subsequent process of mapping which motivates the formal result of the rising pitch question (T_2). The matrix domain represents the space where the two concepts UNKNOWN and TONAL HEIGHT unite (not permanently!) to give birth to a new coherent conceptual domain. If these passages hold,

¹⁹ Theory and practice of singing interestingly divide a “chest voice” from a “head voice” or falsetto. Surely, sharp sounds produced by head voice can be easily sensed putting a hand on our head to feel the vibration, while in chest voice modality we can put one hand on our chest to obtain the same result. That constitutes an experiential base for the metaphor. Moreover, we can refer to the lexical conventionality of this metaphor in the three languages examined, which indicate as “high” (It. “alto”, Mand. “gāo”) or “low” the sound frequencies, often conceptualizing them in the domain of VERTICALITY. These metaphors themselves can suggest to the speaker the overlap of the domains, as it happens even with other well-known metaphors (see e.g. ANGER IS HEAT, Lakoff 1987: 383).

²⁰ In this and other matrices (or matrix domains), the presence of more domains than the ones mentioned is not excluded at all. The main attention here is precisely devoted to that domains which seem to cooperate most to the motivation of linguistic form.

²¹ See Fauconnier 1994; Fauconnier & Turner 1996.

it can be noticed that from an unmarked S , via matrix $(T + T_1)$, we obtain a marked phonological form T . In linear representation, this kind of underlying widespread markedness process can be referred to as: $[S_{\text{unm}} > (T + T_1) > T_{2M}]$.

3.2 Prominence: Intensity variation

In the same fashion, let's start for prominence with an ontological metaphor, namely ARTICULATORY FORCE IS PHYSICAL FORCE, evidently testified by the use we make of phonetic articulatory force, for example, in our everyday quarrels, or to stress the strength of what we are stating. Even in some expressions the metaphor comes out clearly:

(It.) *Urlare contro qualcuno* ("To yell against someone")

(It.) *Gridare (o urlare) in faccia (o addosso) a qualcuno* ("To shout in the face (or to yell at) someone")

(It.) *Parlare sopra qualcuno* ("To speak over someone")²²

Clues of lexical neutralization guide us to the unmarkedness of the cognitive domain (PHYSICAL) FORCE, or rather, to be more precise, the domain (SUFFICIENTLY HIGH) PHYSICAL FORCE.

"Strong/Strength" in neutralization – "Weak/Weakness" not neutralizable

(It.) *È questione di forza/*debolezza* ("It's a matter of strength/*weakness")

(Eng.) *High (or low) strength/*High (or *low) weakness*

(Mand.) *Yǒu duō zhuàng/*ruò?* ("How strong is he?/ How *weak...?"; *zhuàng* = adj. "strong", *ruò* = adj. "weak")

As in the previous case, there is another ontological metaphor to consider in relation to the unmarked starting domain: (+) CONTROL IS (+) FORCE (mainly physical force).

(Eng.) *To force someone to do something*

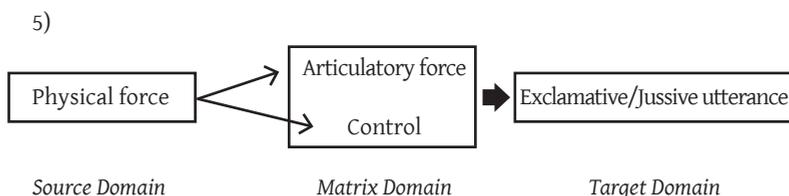
(It.) *Mi ha spinto a farlo* ("He pushed me to do so")

(It.) *Si è sforzato di trattenere la rabbia* ("He struggled strongly to hold his anger")

(It.) *Non voglio forzare nessuno* ("I don't want to force anyone")

²² This example taken from Italian represents also the orientational metaphor CONTROL IS UP.

These two metaphors, combining in a matrix, could be the basis for the motivation of the marked form in some linguistic acts, which we can refer to as exclamative, imperative, and jussive. If we look at Italian, these utterances (or sometimes linguistic acts, for the sake of precision) can display also a dedicated marked (morpho-)syntax: The 2nd person singular imperative forms *Va', Di', Fa', Da'* (instead of *Vai, Dici, Fai, Dai*), together with the marked postverbal or zero subject, represent marked specific constructions that show themselves together with the kind of utterances having the phonetic form we are discussing, to convey this particular pragmatic meaning. To synthesize:



As in 4), the matrix structure could motivate a marked linguistic form T_M starting from an unmarked S . The formula is always the one reminded in the § 3 heading.

3.3 Length variation

At first, we deal with the ontological metaphor (+) EXTENSION/DISTANCE IN TIME IS (+) EXTENSION/DISTANCE IN SPACE²³, a sub-mapping of TIME AS SPACE metaphor, which doesn't need to be discussed here, for it is one of the fundamental metaphors that Cognitive Linguistics fully recognizes as firmly grounded. Instead, we proceed to verify that the cognitive domain of spatial extension is unmarked.

Neutralizable forms:

(It.) *Esteso/Estensione* ("Extended/Extension")

(It.) *Distante/Distanza* ("Distant/Distance")

²³ Based on the highly universal mapping of TIME onto SPACE. See Lakoff & Johnson 2012 [1980]; 1999, Lakoff 1993, Haspelmath 1997.

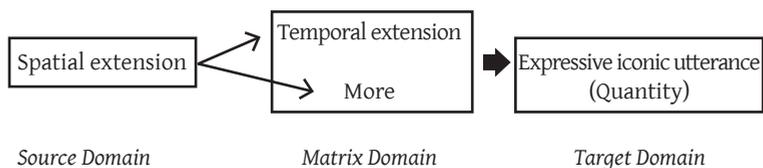
- (It.) *Lungo/Lunghezza* (“Long/Length”)
 (It.) *Ampio/Ampiezza* (“Wide/Width”)
 (Mand.) *Yǒu duō cháng/* duǎn?* (“How long is it?”; *cháng* = adj. “long”, *duǎn* = adj. “short”)

Then, we consider the linear scale metaphor EXTENDED/DISTANT IS MORE, through some examples.

- (Eng.) *This matter is far more complex than I thought*
 (Eng.) *Farther/Further*
 (Eng.) *Widely + adj.*
 (It.) *Superare di gran lunga* (“To largely exceed someone or something”)
 (It.) *La sa (o la dice) lunga* (“He/It knows/tells largely more than he/it is showing”)
 (It.) *Ampiamente migliore (o peggiore)* (“Widely better [or worse]”)
 (Lat.) *Longe* (As “Widely”)
 (Lat.) *Ergo amplius in arte rhetorica nihil est* [*Rhetorica ad Herennium*, end of Book IV] (“So there’s nothing ‘wider’ [more] left to explain about rhetoric”)

Another pragmatic meaning is conveyed by this metaphorical structure. Expressive utterances can partially or entirely extend their duration to be iconic of quantity and size (as in, simplistically, *A biiiiiitig tree!*). We must remember that another quite universal and well-known metaphor fundamentally cooperates to that motivation, even if it is kept on the background: MORE FORM IS MORE CONTENT. Anyway, the schematic representation of our interest is as follows.

6)



Nothing is really different from the previous two cases, and this mapping motivates a marked result starting from an unmarked domain *S*. What we can hypothesize at this point is that the presence of matrices in the structures of motivation can have influence on markedness dynamics, in a radical different way from the chain-like structures.

4 Say ‘*shibboleth*’! Marked phonetic segments and chaining: Type [S_M > T > T_{IM}]

It’s not the purpose of this article to investigate the huge issue of the segments’ cognitive reality or the cognitive ground of the phoneme²⁴. We know from experimental studies by Grimaldi and colleagues (2013) that human brains can access phonetic details if requested by context, and our neuro-cognitive systems can work to categorize segments²⁵. In the Cognitive Linguistics framework, phonemes are conceived as radial categories prototype-centered, and they include more or less central/peripheral allophones as members. Phonemes have conceptual/schematic meaning²⁶, conveyed by the relations allophones constituting the category PHONEME entertain with the prototype.

The biblical episode of the *shibboleth*²⁷ (Judges 12. 6) reveals itself interesting in this perspective. To cut the story short, at the fords of Jordan the Gileadites recognize the fugitive Ephraimites through a test of pronunciation regarding the word “*shibboleth*”. Precisely, the episode deals with a highly marked segment, the [s] belonging to the other linguistic variety spoken by Ephraimites, who can’t pronounce (as traditionally accepted)²⁸ [ʃ] in that word. Conceptualization of this specific segment, and

²⁴ On which, see Albano Leoni 2009. See also Durand & Laks 2002, and Linell 1979.

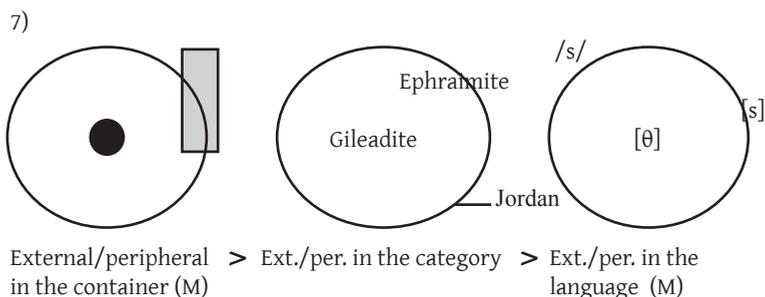
²⁵ As Miglietta-Grimaldi-Calabrese (2013: 288-289) verified, “Two perceptual modes are available for speech perception: a faster phonological (categorical) mode and a slower phonetic (sensory) mode. [...] Our idea is that there is a single neural computation [...] that is ‘sensitive’ to the contrastive/non contrastive status of the sounds as determined by the listeners’ knowledge of the phonological system of their own language [...]. Our results generate the assumption that predictable vowel allophonic alternations (out of context) pattern with phonemic contrasts for auditory perception”.

²⁶ Lakoff 1987: 283: “They [image schemata, including the Container Schema] are concepts that have a directly-understood structures of their own, and they are used metaphorically to structure other complex concepts”.

²⁷ The episode is discussed in the Cognitive Linguistics perspective by Kristiansen (2003), who refers to the allophones [s] and [ʃ], perhaps treating them as parts of a diasystem “Gileadite-Ephraimite”.

²⁸ According to Speiser (1942), the contrast was instead between Ephraimite [s] and Gileadite [θ], because there seems to be no trace of Semitic languages that failed to include both [s] and [ʃ]. The ingenious test put on by the Gileadites proves in itself that they were perfectly aware of the existence of [s] as an allophone, which evokes all the cultural inferences referring to the people of Ephraim. These inferences are tragic for the defeated Ephraimites, who are eventually recognized and killed.

of all the cases of *shibboleth* as a linguistic phenomenon, provides us another example of metaphorical chaining. Two ontological metaphors are involved: A NATURAL CATEGORY IS A CONTAINER²⁹, and A LANGUAGE IS A NATURAL CATEGORY³⁰. In the Gileadites' point of view:



In this mapping, markedness of the peripheral (or external) position of an hypothetical object in relation to the container (therefore, not prototypical) is mapped onto the end of the chaining process, the marked segment [s]. At the intermediate stage, the mapping passes through the peripheral position of the cultural-ethnic element “Ephraimite” in respect to the category which has “Gileadite” as a prototype (for the Gileadites, of course!)³¹. So the result can be expressed by $[S_M > T > T_{1M}]$. If we assume that the perspective can be inverted, what would be transferred to the end of the chain would be the unmarkedness of the same segment instead, as we can easily imagine.

²⁹ See Lakoff & Johnson 1999: 380.

³⁰ One of the most intuitive ontological metaphors, since almost every speaker treats his language(s) as a category, whose elements are (e.g.) English, or they are not members of the category ENGLISH, *et tertium non datur* for many speakers. There are also elements of uncertain status, which speakers peripherally include in their languages, and maybe tend to avoid (in formal contexts, for example).

³¹ There is no mystery that every culture considers itself as the ethnic prototype, and, therefore, every speaker belonging to that category sees himself as prototypical in some degree (at least, comparing himself to ‘foreigners’). Language plays a fundamental role in this categorization, as Cicero reminds us when he refers to a “*Vox Romani generis urbisque propria*” (Cic., *De or.* 3. 12), or when Chinese call their language (including varieties and dialects) “*Hànyü*” (the language of “*Hàn*” dynasty), evidently idealizing their cultural unity under the same category through an as well idealized language.

5. Chains and matrices. A possible role in the motivation of form and markedness

After this brief inquiry, it would be hard to draw definitive conclusions. We can nonetheless make some observations on what presented. First, different metaphorical/schematic structures differently motivate linguistic surface forms and the relative (un-)markedness. Then, since markedness is an important parameter in the discourse on (implicational) universals, trying to understand how it works could be a contribution to understand how universals work. We have also to remember that many directly-emerging cognitive metaphors have a high degree of universality themselves. Thirdly, we can observe that markedness itself could be sensitive to the underlying metaphorical structure motivating it: Matrices seem to provoke the shift from an unmarked domain to a marked linguistic form in certain cases, while chains seem to transfer their (un-)markedness to the formal linguistic result. Lastly, metaphors involved in the matrices mentioned can be of different kinds (orientational, linear scale, ontological), but one of them at least is of the ontological type, none of them (predictably) is a conventional metaphor. What seems to emerge quite clearly is that motivation via cognitive metaphors and schemata contributes to give meaning to apparently non-semantic structures, reducing the gap between grammar and semantics. And the distance seems to be smaller than we usually think.

References

- Albano Leoni, Federico, 2009, *Dei suoni e dei sensi. Il volto fonico delle parole*, Bologna, Il Mulino.
- De Dominicis, Amedeo, 2010, *Intonazione. Una teoria della costituenza delle unità intonative*, Roma, Carocci.
- De Dominicis, Amedeo, 2013, *Fonologie comparate. Suoni e lingue d'Europa, Cina e mondo arabo*, Roma, Carocci.
- Durand, Jacques & Laks, Bernard, 2002, *Phonetics, phonology, and cognition*, Oxford, Oxford University Press.
- Fauconnier, Gilles, 1994, *Mental spaces. Aspects of meaning construction in natural language*, Cambridge, Cambridge University Press.

- Fauconnier, Gilles & Turner, Mark, 1996, "Blending as a central process in grammar", in Goldberg A. E. (ed.), *Conceptual structure, discourse, and language*, Stanford, Center for the Study of Language and Information, 113-131.
- Greenberg, Joseph H., 1963, "Some Universals of Grammar with particular reference to the order of meaningful elements", in Greenberg J. H. (ed.), *Universals of language*, London, MIT Press, 73-113.
- Greenberg, Joseph H., 1966, *Language Universals: With special reference to feature hierarchies*, Berlin, de Gruyter.
- Haspelmath, Martin, 1997, *From space to time. Temporal adverbials in the world's languages*, Munich & Newcastle, Lincom Europa.
- Kristiansen, Gitte, 2003, "How to do things with allophones", in Dirvén R., Frank R., Pütz M. (eds.), *Cognitive models in language and thought*, Berlin & New York, de Gruyter, 69-116.
- Lakoff, George, 1987, *Women, fire, and dangerous things. What categories reveal about the mind*, Chicago, Chicago University Press.
- Lakoff, George, 1993, "The contemporary theory of metaphor", in Ortony A. (ed.), *Metaphor and thought*, Cambridge, Cambridge University Press, 202-251.
- Lakoff, George & Johnson, Mark, 1980, *Metaphors we live by*, Chicago, Chicago University Press.
- Lakoff, George & Johnson, Mark, 1999, *Philosophy in the flesh. The embodied mind and its challenge to Western thought*, New York, Basic Books.
- Lakoff, George & Johnson, Mark, 2012 [1980], *Metafora e vita quotidiana*, ed. & tr. by Violi P., Milano, Bompiani.
- Langacker, Ronald W., 1991, *Foundations of cognitive grammar*, vol. II, *Descriptive application*, Stanford, Stanford University Press.
- Linell, Per, 1979, *Psychological reality in phonology. A theoretical study*, Cambridge, Cambridge University Press.
- Miglietta, Sandra, Grimaldi, Mirko, Calabrese, Andrea, 2013, "Conditioned allophony in speech perception: An ERP study", *Brain & Language*, 126, 285-290 <<<http://dx.doi.org/10.1016/j.bandl.2013.06.001>>>.
- Santoni, Nicola, 2020, "Universali, semantica cognitiva e marcatezza. La motivazione delle forme marcate tra livelli linguistici", in Poli D. & Chiusaroli F. (eds.), *Gli universali e la linguistica. Atti del XLIII Convegno della Società Italiana di Glottologia*, Roma, Il Calamo, 225-235.
- Serra Borneto, Carlo, 2003, "La distribuzione degli ausiliari *haben* e *sein* in tedesco", in Gaeta L. & Luraghi S. (eds.), *Introduzione alla linguistica cognitiva*, Roma, Carocci, 125-146.

- Speiser, Ephraim Avigdor, 1942, "The shibboleth incident (Judges 12: 6)", *Bulletin of the American Schools of Oriental Research*, 85, 10-13.
- Taylor, John R., 2003 [1995], *La categorizzazione linguistica. I prototipi nella teoria del linguaggio*, ed. & tr. by Giannini S., Macerata, Quodlibet. [Taylor, John R., 1995, *Linguistic categorization: Prototypes in linguistic theory*, Oxford, Clarendon Press].
- Yu, Ning, 1998, *The contemporary theory of metaphor: A perspective from Chinese*, Amsterdam & Philadelphia, John Benjamins.
- Yu, Ning, 2009, *From body to meaning in culture: Papers on cognitive semantic studies of Chinese*, Amsterdam & Philadelphia, John Benjamins.

PAOLA VERNILLO*

MIP-IT: Identificazione delle metafore linguistiche in italiano. Una proposta preliminare tra questioni teoriche e problemi pratici

Abstract

L'obiettivo principale del presente contributo è quello di discutere quelli che sono i principali problemi pratici nell'iter di applicazione della MIP, procedura per l'identificazione delle metafore nel discorso, all'italiano. Al di là della ricostruzione dei passaggi cruciali che hanno portato dapprima alla elaborazione del primo modello di annotazione (MIP) e, poi, alla sua estensione (MIPVU), il lavoro propone una serie di linee guida che possano essere utili per la corretta e rigorosa identificazione delle metafore linguistiche in italiano (MIP-IT). Tali norme, pur ponendosi in linea con quanto già stabilito dalla MIP e, in parte, dalla MIPVU, tengono conto del particolare contesto di applicazione e, dunque, partono dalla considerazione delle peculiarità grammaticali e linguistiche dell'italiano. A corollario di quanto problematizzato nel corso della riflessione teorica qui proposta, presentiamo un esempio di applicazione pratica.

Parole chiave: MIP, MIPVU, Metafore linguistiche, metafore concettuali

This contribution aims to discuss the main practical problems in applying MIP, a metaphor identification procedure in discourse, to the Italian language. In addition to describing the crucial steps behind the development of the first annotation model (MIP) and its extension (MIPVU), this work provides a set of rigorous and proper guidelines useful for identifying linguistic metaphors in Italian (MIP-IT). While aligning with MIP and, albeit partially, MIPVU, these guidelines consider the specific application context and are therefore based on the grammatical and linguistic peculiarities of Italian. Moreover, this work presents an example of the practical application of MIP-IT as a corollary to the reported issues faced during the theoretical reflection proposed herein.

Keywords: MIP, MIPVU, linguistic metaphors, conceptual metaphors

* Paola Vernillo, Alma Mater Studiorum - Università di Bologna, paola.vernillo@unibo.it.

1. Introduzione

Strumento di comparazione, anomalia grammaticale, semantica e concettuale (Beardsley 1962; Loewenberg 1975; Levin 1977), ma anche imprescindibile mezzo di intermediazione tra domini concettuali distinti (Black 1954, 1962, 1979; Ortony 1979; Ricoeur 2010 [1975]): la storia della metafora e della sua feconda carriera si dipana nel tempo in modi decisamente sorprendenti. Ciò che poco sorprende invece è il sapere che attualmente la metafora rappresenta una delle aree di indagine più ampiamente investigate nell'ambito degli studi della Linguistica, in generale, e della Semantica, in particolare.

Ponendosi come un fenomeno poliedrico, indagabile sia sul piano del linguaggio sia sul piano concettuale, con ricadute significative soprattutto negli studi di Semantica e Pragmatica (Searle 1993), stupisce però che gli aspetti precipuamente metodologici per una sua corretta identificazione all'interno del discorso siano stati per lungo tempo trascurati (Steen et al. 2010). La parzialità dell'intuizione su cui si sono spesso basati i giudizi di singoli esperti, in assenza di criteri di identificazione espliciti e condivisi, ha comportato non soltanto un certo grado di variabilità nei risultati delle analisi condotte ma anche un quantitativo di vaghezza nelle operazioni con cui si è tentato di determinare ciò che è effettivamente metaforico (Cienki 2008: 254; Pragglejaz Group 2007: 2).

Come si fa dunque a discernere ciò che è metafora da ciò che non lo è? E come si fa a produrre delle pratiche di annotazione che siano fondate su criteri di individuazione stabili e soprattutto coerenti?

L'obiettivo primario del presente studio, che pure è sorretto da un solido assetto teorico nonché da una concezione della metafora come duplice figura del pensiero e del discorso (Lakoff & Johnson 1980), è quello di discutere tale fenomeno da un punto di vista primariamente tecnico, pratico e metodologico. Più nello specifico, ciò che ci interessa approfondire in queste pagine è l'applicabilità all'italiano¹ della MIP (Metaphor Identification Procedure: Pragglejaz 2007) e MIPVU (Steen et al. 2010), protocolli di annotazione me-

¹ Sulla base dello spoglio della letteratura esistente e del confronto avuto con alcuni dei membri storici del gruppo MIP/MIPVU (in particolare, Alan Cienki, Tina Krennmayr ed Elena Semino), posso affermare che, al momento, non esistono altri casi di applicazioni del protocollo all'italiano. Per una panoramica sull'applicazione ad altre lingue si veda Nacey et al. (2019).

taforica ormai ben collaudati per l'inglese e già ampiamente testati anche su altre lingue, tra cui francese, tedesco, olandese e norvegese (Nacey et al. 2019). Lo studio partirà dalla considerazione in chiave critica delle problematiche specificatamente legate ad aspetti idiosincratici, sia sul piano lessicale che grammaticale, dell'italiano e servirà ad elaborare delle linee guida lingua-specifiche che permettano una corretta e allo stesso tempo flessibile estensione delle suddette procedure (MIP-IT). La proposta di estensione sarà testata su un breve testo selezionato da un mini-corpus, creato sulla base della estrazione semi-automatica di video dalla piattaforma TikTok e contenente trascrizioni di parlato spontaneo prodotto da soggetti italofofoni che hanno sofferto di anoressia nervosa (Donati et al. 2023).

Nel Paragrafo 2, presentiamo una breve introduzione al concetto di metafora concettuale, mentre una più ampia prospettiva sulla genesi e sulla articolazione del protocollo MIP è fornita nei paragrafi 3, 3.1 e 3.2. Nella sezione successiva (par. 4) segue una breve discussione sulla complessità e varietà delle questioni operative legate all'applicazione della procedura all'italiano. Il paragrafo 5 rende esplicite le linee guida che saranno adottate nell'ambito della nostra proposta di estensione e discute un esempio di analisi. Le conclusioni sono presentate nel par. 6.

2. La Teoria della Metafora Concettuale

Tutto ciò che sappiamo sulla metafora è il prodotto di un instancabile processo di ricerca e riflessione iniziato quando Aristotele per primo prova a definirne i confini e con essi lo statuto tropico. Nel corso degli anni, la concezione aristotelica è sottoposta a una duplice lettura (Black 1954; Cameron 2003; Charteris-Black 2000): da un lato, la teoria della sostituzione, secondo cui l'espressione metaforica sarebbe usata al posto di una equivalente espressione letterale per trasferire il nome proprio di una cosa ad un'altra (A è B); dall'altro, la teoria del confronto, in accordo alla quale la metafora si porrebbe come una sorta di similitudine abbreviata (A è come B) sorretta da un rapporto analogico. Punto di congiunzione tra le due interpretazioni è l'idea secondo cui fare buoni traslati è saper vedere ciò che è somigliante (Poet. 22, 59 a, 5-9). È questa stessa idea, tuttavia, a diventare anche il bersaglio principale delle critiche mosse dalle teorizzazioni successive (Charteris-Black 2004), prima fra tutte la concezione interazionista di Max Black (1954, 1962), secondo cui la

metafora non si limiterebbe a stabilire a priori una somiglianza oggettiva tra cose, ma svolgerebbe una vera e propria funzione euristica, di creazione di qualcosa non già esistente e non egualmente esprimibile attraverso il ricorso al linguaggio letterale. È probabilmente Black² a liberare il passaggio a una nuova visione, in chiave primariamente cognitiva, della metafora, affrancandola dal ruolo di mero segnale di scarto tra uso poetico e linguaggio ordinario, e sganciandola dal piano dell'analisi puramente linguistica. Il passaggio verso un approccio più concettivo allo studio della metafora si compie però con la pubblicazione, nel 1980, del libro *Metaphors we live by* scritto da George Lakoff e Mark Johnson. Tale volume costituisce una pietra miliare nella storia degli studi semantici e una tappa fondamentale nel percorso di espansione della Linguistica Cognitiva. All'interno di questa cornice teorica, la cosiddetta Teoria della Metafora Concettuale (TMC), la metafora smette finalmente i panni del mero abbellimento discorsivo (Searle 1993), per essere rivestita di una identità nuova, primariamente concettuale (Lakoff & Johnson 1980: 4):

Metaphor is for most people a device of the poetic imagination and the rhetorical flourish—a matter of extraordinary rather than ordinary language. Moreover, metaphor is typically viewed as characteristic of language alone, a matter of words rather than thought or action. For this reason, most people think they can get along perfectly well without metaphor. We have found, on the contrary, that metaphor is pervasive in everyday life, not just in language but in thought and action. Our ordinary conceptual system, in terms of which we both think and act, is fundamentally metaphorical in nature.

La metafora, dunque, è concepita come un fenomeno sistematico e non singolare, cognitivo prima che linguistico, attraverso cui strutturiamo e rappresentiamo un dominio semantico nei termini di un altro. Una delle conseguenze più significative di questo cambio di assetto paradigmatico riguarda probabilmente la considerazione delle diverse modalità attraverso cui esprimiamo concetti metaforici, i quali, lungi dall'essere esplicitabili sul piano strettamente verbale, sono anzi esprimibili attraverso gesti (Cienki 1998; Cienki & Muller 2008; McNeill 1992;), immagini (Forceville & Urios-Aparisi 2009) e, più in generale, qualunque altro tipo di codice comunicativo o artistico.

² Fondamentali sono stati lavori pionieristici di Richards (1936).

Uno dei pilastri su cui si regge l'impalcatura teorica della TMC è la chiara e netta distinzione tra due categorie di oggetti metaforici: da un lato le metafore concettuali, dall'altro le metafore linguistiche. Più nello specifico, la TMC definisce le metafore linguistiche nei termini di espressioni che rappresentano, sul piano del linguaggio verbale, strutture concettuali metaforiche profonde. Per esempio, le metafore linguistiche "Mi trascino in una esistenza priva di significato" o "La ballerina si è spinta oltre i propri limiti", esplicitano linguisticamente una struttura metaforica che pone in essere una relazione tra agito e movimento nello spazio. Entrambe le espressioni sono infatti basate su una concettualizzazione dell'agito personale nei termini di un movimento autoindotto lungo un percorso spazialmente esteso (Lakoff 1993). In altre parole, nell'ottica della TMC, queste stesse due espressioni metaforiche esplicitano, seppur con scelte lessicali differenti, lo stesso tipo di struttura metaforica, potendo essere entrambe interpretate come il riflesso linguistico della metafora concettuale SELF-INITIATED ACTIONS ARE SELF-PROPELLED MOTIONS. Lo schema di questa metafora segue la formulazione canonica "A è B", in accordo alla quale un dominio astratto (o target) A è mappato su un dominio concreto (o source) B, dal quale si estraggono contenuti informativi che sono sistematicamente e unidirezionalmente trasferiti per la modellazione del dominio A. In questo caso specifico, dunque, costruiamo il dominio astratto dell'agito personale sulla base del trasferimento di una selezione di tasselli informativi propri del dominio concreto del movimento.

3. Metodi per l'identificazione delle metafore nel discorso

La pubblicazione di *Metaphors we live by* ha rappresentato un punto di snodo storico la cui portata ha avuto effetti più o meno immediati ma certamente duraturi nell'ambito degli studi di linguistica (Gibbs 2008). L'imponenza concettuale di questo framework teorico si misura non solo nelle innumerevoli applicazioni in studi appartenenti a campi di ricerca anche molto diversi tra loro (tra le varie, Linguistica, Psicologia, Filosofia, Analisi del discorso, Antropologia, Neuroscienze e Informatica) ma anche e soprattutto nella quantità di approcci che si sono posti in linea di continuità o di rottura con lo stesso (Geeraerts 2010; Grady 1997; Lakoff 1993; Steen 1999, 2007). Ciò che tuttavia per lungo tempo è sembrato

mancare all'interno di questo articolato assetto architettonico è stato l'interesse precipuo per lo sviluppo di protocolli o metodi per una identificazione sistematica, affidabile e replicabile delle metafore linguistiche nel discorso (Semino 2008). L'assenza di tali linee guida ha fatto sì che per parecchio tempo ricercatori e studiosi nel campo della metafora si siano affidati alle proprie intuizioni individuali (Kövecses 2011) e, in assenza di criteri formalmente condivisi, abbiano valutato in modo tendenzialmente discrezionale il confine tra letterale e non letterale (Gibbs & Colston 2012). Tale variabilità ha necessariamente impattato non solo sulla affidabilità dei risultati ottenuti nei singoli studi, sia in termini quantitativi che qualitativi, ma anche sulla comparabilità, in chiave intra e interlinguistica, degli esiti a cui son giunte le diverse ricerche nel campo metaforologico (Cameron 2003; Pragglejaz 2007; Semino, Heywood, & Short 2004; Steen et al. 2010). Tuttavia, negli ultimi anni, qualcosa sembra esser cambiato e l'applicazione di un metodo rigorosamente scientifico per l'identificazione di metafore nel discorso non sembra più rappresentare un problema così cogente (Steen 2017), potendo i ricercatori contare su una letteratura ampia nonché su una quantità di modelli validi e ben collaudati su ampi set di dati (Cameron 1999; Cameron e Maslen 2010; Charteris-Black 2004; Littlemore et al. 2013; Schmitt 2005). Nei prossimi paragrafi, non ci soffermeremo sulla descrizione dei principali protocolli ad ora a disposizione, ma proveremo a tracciare il profilo di uno tra i più imponenti e apprezzati sistemi per l'identificazione di metafore nel testo (MIP), cercando anche di delineare similitudini e differenze con l'estensione che ne è stata successivamente proposta (MIPVU). Tale ricostruzione è funzionale al tentativo di ampliare il bacino di applicazione di questi stessi protocolli, che qui testeremo su una lingua come l'italiano per cui, di fatto, non esistono linee guida precise e chiare che possano regolamentarne l'uso e assicurarne una loro corretta applicazione.

3.1 MIP: *Metaphor Identification Procedure*

Nel 2007 la rivista *Metaphor & Symbol*, punto di riferimento nell'ambito degli studi sulla metafora, pubblica *MIP: A method for identifying metaphorically used words in discourse*, studio nato dallo sforzo congiunto di dieci ricercatori uniti dal comune obiettivo di sviluppare un protocollo chiaro, affidabile, flessibile e replicabile per l'identificazione delle metafore nel discorso (Pragglejaz 2007). Il team, in realtà, inizia a collaborare stabilmente già a partire dai pri-

mi anni 2000, anno in cui Peter Crisp, Raymond Gibbs, Alice Deignan, Graham Low, Gerard Steen, Lynne Cameron, Elena Semino, Joe Grady, Alan Cienki e Zoltan Kovecses si uniscono a formare il cosiddetto gruppo Pragglejaz, acronimo nato proprio dalla combinazione delle iniziali dei componenti originari (Nacey et al. 2019; Semino 2019; Steen 2017). Per quel che concerne la procedura, la Metaphor Identification Procedure (in breve, MIP) impone il rispetto di una serie articolata di punti funzionali alla identificazione delle cosiddette metafore indirette, ossia unità lessicali³ utilizzate metaforicamente all'interno di un testo, sia esso scritto o orale. A tal proposito, è importante sottolineare che tale metodo di analisi si focalizza sulle sole metafore linguistiche e lascia di fatto al di fuori del task interpretativo fornito dal protocollo l'individuazione e l'analisi delle metafore concettuali⁴. L'interesse precipuo della MIP, dunque, è quello di annotare in modo rigoroso e coerente i soli dati linguistici. Più nello specifico, la metodologia di analisi linguistica suggerita dalla MIP si fonda sul confronto tra il significato contestuale e il significato base di ciascuna unità lessicale in cui è possibile segmentare il testo in analisi (Pragglejaz 2007: 3):

1. Read the entire text–discourse to establish a general understanding of the Meaning.
2. Determine the lexical units in the text–discourse.
3. (a) For each lexical unit in the text, establish its meaning in context, that is, how it applies to an entity, relation, or attribute in the situation evoked by the text (contextual meaning). Take into account what comes before and after the lexical unit.

(b) For each lexical unit, determine if it has a more basic contemporary meaning in other contexts than the one in the given context. For our purposes, basic meanings tend to be:

³ In linea generale, salvo eccezioni, le unità lessicali coincidono con le unità grafiche, le parole ortografiche.

⁴ È bene, tuttavia, ricordare che, pur non fornendo strumenti pratici per l'individuazione di metafore concettuali nel discorso, la MIP (ma anche la sua successiva estensione in MIPVU) nasce e si sviluppa all'interno di un framework teorico che adotta un approccio fortemente cognitivo all'esplorazione del fenomeno metaforico (Lakoff & Johnson 1980), in virtù del quale le metafore linguistiche sono considerate come il riflesso, sul piano verbale, di un fenomeno di mappatura che è primariamente concettuale.

More concrete.

What they evoke is easier to imagine, see, hear, feel, smell, and taste.

Related to bodily action.

More precise (as opposed to vague).

Historically older.

Basic meanings are not necessarily the most frequent meanings of the lexical unit.

(c) If the lexical unit has a more basic current–contemporary meaning in other contexts than the given context, decide whether the contextual meaning contrasts with the basic meaning but can be understood in comparison with it.

4. If yes, mark the lexical unit as metaphorical.

Uno degli aspetti distintivi della metodologia di indagine su cui si fonda la MIP è l'uso sistematico di risorse lessicali con cui stabilire eventuali differenze tra le tipologie di significato sopra menzionate. Per l'inglese, la risorsa lessicale scelta come fonte informativa primaria è il *Macmillan English Dictionary for Advanced Learners* (Rundell & Fox 2002), un dizionario di inglese contemporaneo che è basato su un corpus contenente circa 220 milioni di parole. Tuttavia, oltre al dizionario *Macmillan*, il protocollo MIP raccomanda, come fonte aggiuntiva di informazione etimologica, la consultazione dello *Shorter Oxford English Dictionary on Historical Principles* (SOEDHP: Little et al. 1973).

Per chiarire più nel dettaglio il funzionamento della procedura, si consideri il seguente esempio (Pragglejaz 2007):

(1) For years Sonia Gandhi has struggled to convince Indians [...]

Conformemente al protocollo MIP, il testo in (1) andrà internamente segmentato in unità lessicali:

(1) / For / years / Sonia Gandhi / has / **struggled** / to / convince / Indians [...]

Una volta individuate le unità lessicali di cui si compone il testo, la consultazione del dizionario ci permetterà di valutare la presenza dell'eventuale scarto tra significato contestuale e significato base di ciascuna unità lessicale. In

questa sede, per necessità di semplificazione, circoscriveremo l'applicazione del protocollo alla sola parola *struggled*⁵:

- **Significato contestuale:** indica sforzo o difficoltà nel raggiungere un certo obiettivo.
- **Significato base:** indica l'usare forza fisica contro qualcuno o qualcosa.

Come si evince dal confronto tra le due tipologie di significato, il significato contestuale si pone come una sorta di significato indiretto (Gibbs 1994) dal valore metaforico:

- **Significato contestuale vs significato base:** il significato contestuale diverge dal significato base poiché la parola *struggled* si riferisce a uno sforzo o difficoltà da intendere sì in senso figurato (o astratto), ma concepibile nei termini di uno sforzo o difficoltà prettamente fisici⁶.

Al termine di questo iter procedurale, come si suggerisce anche in Pragglejaz (2007), l'esito positivo del task di comparazione non può che portarci ad indicare come metaforico l'uso che contestualmente è fatto dell'unità lessicale in analisi.

Come anticipato, sia per necessità di semplificazione sia per motivi di spazio, abbiamo deciso di illustrare una piccola parte del protocollo di applicazione, annotando la sola parola *struggled*. Tuttavia, tale taglio non ci ha permesso qui di mostrare le eventuali difficoltà in cui si potrebbe incappare durante un esercizio di annotazione come quello definito dalla MIP. Difatti, stabilire il significato base di un'espressione non è affatto una operazione facile e potrebbero essere molti i casi in cui gli annotatori non riescano a giungere a una soluzione condivisa e, soprattutto, definitiva, in tempi più o meno rapidi. Per

⁵ È interessante notare che la MIP ignora la distinzione tra classi di parole e dunque tutte quelle informazioni che sono associabili all'appartenenza di un item a una specifica parte del discorso (PoS). In altri termini un nome come *squirrel* e un verbo come *to squirrel* non sono trattati come lessemi distinti. I creatori della MIP motivano tale decisione sulla base della necessità di preservare la relazione metaforica tra le due parole (Pragglejaz 2007: 28).

⁶ Lo scarto tra il significato contestuale e il significato base è in questo caso rinforzato anche dal fatto che lo stesso dizionario etimologico *Shorter Oxford Dictionary on Historical Principles* indica il significato base come già storicamente precedente.

esempio, potrebbe certamente essere meno complesso individuare il significato base di una parola contenuto come un nome piuttosto che di una parola grammaticale come una congiunzione. Allo stesso tempo potrebbe essere più semplice assegnare un significato base a un verbo azionale, dalla semantica primariamente concreta, piuttosto che a un verbo con funzione ausiliaria dal significato più astratto. All'interno del protocollo MIP, tali problemi non sono lasciati al caso ma sono risolti considerando preposizioni (a parte rari casi), congiunzioni, verbi ausiliari, pronomi come elementi dotati di significato astratto e dunque non bollabili come elementi utilizzati metaforicamente (Pragglejaz 2007: 29). Ovviamente, questo tipo di difficoltà decisionali e applicative, così come tutte le criticità legate alla segmentabilità del testo in unità lessicali, alla prossimità tra metafore e metonimie, nonché alla consultazione delle fonti lessicografiche, sono affrontate con rigore scientifico. Ciò significa che l'iter procedurale proposto dalla MIP prevede anche una validazione delle annotazioni e dei dati da queste ricavati, consistente in una fase di discussione aperta tra i singoli annotatori e una fase di misurazione dell'accordo tra gli stessi (ad es., calcolo del coefficiente statistico *Kappa di Cohen*).

3.2 Dalla MIP alla MIPVU

Una magistrale ricostruzione temporale del passaggio dalla MIP alla MIPVU⁷ si deve a Susan Nacey (2013), la quale in *Metaphors in Learner English* descrive la MIPVU non nei termini di una semplice estensione della precedente procedura di identificazione MIP ma come il risultato di un vero e proprio processo di reincarnazione che trova il suo culmine nel 2010, anno in cui Gerard Steen e il suo team di collaboratrici olandesi danno alle stampe *A Method for Linguistic Metaphor Identification. From MIP to MIPVU*. Questo lavoro si pone come il prodotto finale di un progetto di ricerca in cui la MIP è sistematicamente applicata a un corpus di testi differenziabili per registro (narrativa, discorso accademico, discorso dialogico e giornalistico: Steen 2002, 2017; Steen et al. 2010). I punti di divergenza parziale tra i due protocolli di annotazione sarebbero essenzialmente quattro: l'individuazione delle unità lessicali, la ridefini-

⁷ L'abbreviazione VU si riferisce alla Vrije Universiteit, l'Università olandese, con base ad Amsterdam, in cui è stato condotto il lavoro di revisione del protocollo MIP (Steen et al. 2010).

zione del senso base, l'allargamento del bacino di metaforicità con l'aggiunta di metafore dirette e implicite (oltre alle cosiddette metafore indirette), l'aggiunta di tre nuove etichette di annotazione (WIDLII: When In Doubt, Leave It In; DFMA: Discarded For Metaphor Analysis; PP: Possible Personification). Il primo punto differenziale riguarda le modalità di scomposizione del testo in singole unità testuali e influenza necessariamente il protocollo sin dal suo avvio (Step 1). In linea generale, sia la MIP che la MIPVU tendono a identificare le unità lessicali con le parole ortografiche⁸. Tuttavia, mentre la MIP esclude dall'equazione verbi sintagmatici (ing. *phrasal verbs*), composti e polirematiche (ing. *multiword expressions* e *polywords*)⁹, la MIPVU aggiunge alla precedente operazione le polirematiche e ne sottrae alcuni specifici tipi di nomi propri (Nacey 2013). Per quel che riguarda il secondo punto, la definizione di ciò che sia da intendersi per significato base di una certa parola (Step 3b) subisce una sorta di restringimento semantico in virtù del quale l'aspetto etimologico perde di significatività nel processo decisionale. Tale scelta sarebbe motivata dal fatto che: a) il parlante non avrebbe reale coscienza della etimologia della parola utilizzata (Steen et al. 2010); b) con l'esclusione del fattore etimologico si semplificherebbe enormemente il protocollo; c) il significato più concreto sarebbe già quello storicamente precedente (vedi esperienzialismo: Lakoff & Johnson 1980). A marcare in maniera significativa l'allontanamento delle due procedure ci sarebbe però l'allargamento del cosiddetto bacino di metaforicità. Difatti, mentre la MIP si fonda sulla identificazione delle sole metafore indirette (par. 3.1.), la MIPVU suggerisce di estendere l'indagine sia alle cosiddette metafore dirette sia alle metafore implicite¹⁰. Più nello specifico, le metafore dirette sono espressioni che pon-

⁸ Parole ortograficamente distinte, dunque separate da spazi, rappresentano unità lessicali distinte (Steen 2017).

⁹ All'interno del protocollo MIP/MIPVU, i cosiddetti *idioms* (ad es., ing. *spill the beans*; ita. "vuotare il sacco"; traduzione letterale "rovescia il fagiolo") non sono considerati come unità lessicali unitarie ma come espressioni segmentabili in componenti essenziali (ad es., / spill / the / beans /).

¹⁰ Nel caso delle metafore dirette e implicite non sarebbe corretto parlare di *Metaphor Used Words*. La lieve riformulazione terminologia adottata dalla MIPVU, tenuto conto di questa problematica, preferisce riferirsi alle tre diverse tipologie metaforiche mediante l'espressione *Metaphor-Related Word*, (o anche MRW).

gono in essere una comparazione tra due domini senza però comportare per questo un contrasto tra significato contestuale e significato base dell'espressione. In questo riassetto tassonomico, la MIPVU, dunque, arriva a inglobare anche le similitudini (in inglese, gli indicatori linguistici sarebbero rappresentati da espressioni come *like* e *as*: Steen 2008) e le analogie. Si consideri il seguente esempio (Steen et al. 2010: 57):

(2) For many years Thompson lived in New York in his apartment at the Chelsea Hotel. From there, **like a buzzard in its eyrie**, he would make forays round the US and abroad (...)

Espressioni come quelle segnalate in grassetto, non sono utilizzate metaforicamente ma attivano una equazione diretta. L'esempio in (2), difatti, si fonda sull'esplicitazione linguistica (*like*) della comparazione tra i due domini semantici (da un lato, un agente umano; dall'altro, un rapace) implicati nel processo di trasposizione simbolica (*Thompson* e *buzzard*).

Con l'espressione metafore implicite, invece, non ci si riferisce a strutture linguistiche dotate di metaforicità ma a strumenti di coesione testuale (ad es., anafora) a loro volta riferibili a porzioni di testo effettivamente utilizzate in modo non letterale ma figurato (Steen et al. 2010: 122):

(3) Fortunately, there is a single antidote effective against both these myths; and **that** is to start all over again with the most decisive source of Darwin's new identity, on the voyage, as a committed man of science: his zealous discipleship of Charles Lyell's (1797–1875) views in geology (including biogeography and ecology).

Nell'esempio (3) il pronome *that* attiva una strategia di implicazione metaforica, sulla base di un processo di sostituzione (*that* si riferisce qui ad *antidote*). La metafora implicita, dunque, è qui attivata dal fatto di aver utilizzato un elemento grammaticale che non è di per sé metaforico ma che si riferisce ad una porzione del testo precedente che è effettivamente utilizzata in modo non letterale.

A chiusura di questa operazione di ricostruzione dei punti di divergenza tra i due protocolli, portiamo la discussione delle tre etichette supplementari: WIDLII ('When In Doubt, Leave It In'), DFMA ('Discarded For Metaphor Analysis') e PP ('Possible Personification'). Il primo acronimo, WIDLII, è funzionale alla risoluzione di quei casi in cui non si sia in grado di rispondere in

modo certo e definitivo alle perplessità sulla metaforicità di una certa unità lessicale. In altri termini, in tutti quei casi in cui la binarietà del giudizio non sia praticabile (metaforico/non metaforico), la MIPVU suggerisce di categorizzare l'unità come una sorta di caso limite che pur nella ambiguità del suo statuto semantico si decide di tenere all'interno della raccolta. Il secondo acronimo, DFMA, è invece riservato a tutti quei contesti in cui non sia effettivamente possibile applicare la metodologia di identificazione (ad es., nelle disfluenze discorsive). Infine, il terzo ed ultimo acronimo, PP, serve a indicare la presenza di un particolare tipo di processo di comparazione, ossia i casi di personificazione, basati sull'attribuzione di proprietà umane ad entità che umane non sono (per es., questo vento non ha intenzione di placarsi). L'esito di questo processo di trasformazione della MIP in MIPVU può essere esemplificato dalla revisione delle linee guida raccomandate per la corretta applicazione della nuova procedura di identificazione delle metafore (MIPVU) e che, per esaustività, riportiamo qui di seguito (Steen et al. 2010: 25-26):

1. Find metaphor-related words (MRWs) by examining the text on a word-byword basis.
2. When a word is used indirectly and that use may potentially be explained by some form of cross-domain mapping from a more basic meaning of that word, mark the word as metaphorically used (MRW).
3. When a word is used directly and its use may potentially be explained by some form of cross-domain mapping to a more basic referent or topic in the text, mark the word as direct metaphor (MRW, direct).
4. When words are used for the purpose of lexico-grammatical substitution, such as third person personal pronouns, or when ellipsis occurs where words may be seen as missing, as in some forms of co-ordination, and when a direct or indirect meaning is conveyed by those substitutions or ellipses that may potentially be explained by some form of cross-domain mapping from a more basic meaning, referent, or topic, insert a code for implicit metaphor (MRW, implicit).
5. When a word functions as a signal that a cross-domain mapping may be at play, mark it as a metaphor flag (MFlag).
6. When a word is a new-formation coined, examine the distinct words that are its independent parts according to steps 2 through 5.

4. MIP-IT: Identificazione delle metafore in italiano

Questo studio si propone di testare l'applicabilità della procedura per l'identificazione delle metafore, MIP, sull'italiano (MIP-IT). In accordo alla categorizzazione proposta dal gruppo Pragglejaz (2007), l'annotazione qui condotta sarà focalizzata sulle sole metafore indirette e, dunque, lascerà al di fuori del raggio di interesse sia le metafore dirette che le metafore implicite. L'importanza di tale indagine si misura su un duplice piano, teorico e pratico. Da un punto di vista teorico, infatti, essa ci permette di condurre una disamina critica di ambo potenzialità e limiti applicativi della MIP sull'italiano, lingua per cui, diversamente da quanto accade, per esempio, per il francese (Reijnierse 2019), non esistono né un protocollo né delle linee guida chiare e trasparenti. L'esito atteso è dunque l'elaborazione di norme specifiche attraverso cui definire un modello di identificazione che sia valido anche all'interno del nostro pur complesso quadro linguistico. Da un punto di vista pratico, questa stessa indagine ci permette di valutare su dati testuali reali gli ostacoli concretamente riscontrabili, ma anche le possibili modalità di risoluzione, in questa operazione di ricerca della metaforicità attraverso l'impiego del protocollo MIP. Nei sottoparagrafi successivi, discuteremo una serie di caratteristiche lessicali e grammaticali dell'italiano che possono condizionare o, addirittura, complicare il task di identificazione. Particolare attenzione sarà dedicata dunque a quegli elementi che influenzano l'operazione di scomposizione del testo in unità lessicali: polirematiche, composti, verbi sintagmatici e verbi pronominali.

4.1 MIP-IT e costruzioni complesse: il caso delle polirematiche

Le polirematiche sono combinazioni di parole strutturalmente coese e semanticamente unitarie¹¹ che gli studiosi tradizionalmente collocano a metà strada tra morfologia e sintassi (De Mauro 1999; Masini 2011; Simone 2006; Voghera 2004). Le varietà di denominazioni che in letteratura sono state adottate per riferirsi alle polirematiche, anche note come lessemi complessi

¹¹ In genere, si sostiene che il significato delle espressioni polirematiche non sia compositivo, dal momento che esso non corrisponderebbe alla semplice somma dei significati delle parole parte del costruito. In tal senso, da un punto di vista semantico le polirematiche si comporterebbero non come unità nate dalla giustapposizione di parole distinte ma come una unica unità lessicale.

(De Mauro & Voghera 1996), parole complesse (Simone 2008), unità lessicali superiori (Dardano 1978), parole sintagmatiche o espressioni multiparola (Masini 2009, 2012), sono perfettamente simmetriche alla eterogeneità delle forme linguistiche che queste stesse marche terminologiche designano. Le polirematiche, infatti, rappresentano una classe mista (espressioni idiomatiche, formule conversazionali, proverbi, ma anche citazioni o interi testi) al cui interno si collocano strutture ascrivibili a diverse categorie lessicali¹² (Masini 2009): nomi (*macchina da scrivere*), aggettivi (*alla mano*), avverbi (*avanti e indietro*), verbi (*tirare su*), ma anche pronomi (*il tal dei tali*), preposizioni (*di fronte a*), congiunzioni (*in quanto*) e interiezioni (*mamma mia*). Come è ovvio aspettarsi, la non omogeneità dei fenomeni fin qui descritti e l'assenza, per quel che concerne l'italiano, di una lista che sia davvero condivisa e completa (Masini et al. 2020; Simone & Piuanno 2017) hanno un impatto negativamente significativo sulle decisioni da prendere per la loro segmentazione all'interno di un testo (Step 1 MIP). Come comportarsi dunque con le espressioni polirematiche? Dovremmo considerarle come composte da unità lessicali multiple ma separabili o piuttosto come singole unità lessicali? Sia la MIP che la MIP-VU raccomandano di trattare le polirematiche (eng. *polywords*) come singole unità lessicali (Nacey et al. 2019; Praggeljaz Group 2007), motivando tale scelta sulla base della disponibilità, per quel che concerne i dati inglesi, di liste di polirematiche liberamente consultabili sul sito del British National Corpus¹³ (Steen et al. 2010). Tuttavia, questa non sembra essere una via immediatamente praticabile per quegli annotatori che intendano occuparsi di dati di italiano, contesto in cui, anche a livello prettamente lessicografico, non sempre si registrano atteggiamenti uniformi nel trattamento di queste stesse espressioni. La locuzione *in quanto*, per esempio, è lemmatizzata¹⁴ e resa consultabile come entrata a parte nel GRADIT (Grande Dizionario Italiano: De Mauro 1999), dove è definita come: 1) locuzione congiuntiva, in proposizioni causali

¹² Interessanti, in tale contesto, si rivelano i lavori per la costruzione della risorsa PARSEME: <https://typo.uni-konstanz.de/parseme/>

¹³ www.natcorp.ox.ac.uk.

¹⁴ In linea generale, le polirematiche sono lemmatizzate e definite sotto il loro primo elemento pieno. Tuttavia, un piccolo gruppo di polirematiche figura anche a lemma autonomo (questo è il caso, per esempio, di locuzioni latine o di lingua straniera).

con verbo all'indicativo, col significato di “perché, dal momento che”; 2) locuzione avverbiale col significato di “in qualità di, come”. Tuttavia, a questa stessa locuzione è riservato un trattamento diverso ne *Il vocabolario Treccani*, che decide di non creare una entrata a parte ma di fornire una definizione più dettagliata alla voce “quanto”, dove compare come accezione particolare del lemma generale col valore di “locuzione con valore avverbiale e congiuntivo, per introdurre una proposizione che limita il concetto espresso nella principale”. Come è evidente, sulla base dell'indicazione fornita dall'una o dall'altra fonte lessicografica, saremmo portati a segmentare queste espressioni in modo diametralmente opposto. Considerato questo stato di cose, la difficoltà di demarcazione delle unità polirematiche rende necessaria l'adozione di linee guida che siano specificatamente calibrate sull'italiano e che quindi siano supportate da valutazioni che tengano conto sia del contesto scientifico di implementazione del protocollo MIP sia della necessità di trovare soluzioni che siano anzitutto funzionali alla buona riuscita del nostro task di analisi. È per tale ordine di motivi che abbiamo deciso di adottare un approccio pragmatico al trattamento delle espressioni polirematiche, in virtù del quale la decidibilità sulla segmentazione dei lessemi complessi venga a dipendere dalla caratterizzazione che ne è data all'interno della fonte lessicografica che qui abbiamo scelto come preferenziale, vale a dire il GRADIT. In particolare, valuteremo come singole unità lessicali quelle espressioni che siano state riportate a lemma e come riducibili in unità lessicali minori quelle espressioni per cui, al contrario, non sia stata resa disponibile questa opzione¹⁵.

4.2 MIP-IT e costruzioni lessicali complesse: il caso dei composti

Al pari delle polirematiche¹⁶, i composti sono trattabili come formazioni lessicali collocabili a metà strada tra il piano morfologico e quello sintattico. Tali strutture possono nascere dall'unione di due o più parole (ad es., combinazioni di due nomi come in *pesce spada*), parti di parole (ad es., combinazioni

¹⁵ Per motivi pratici, la nostra procedura si discosta da quanto raccomandato in studi precedenti sul francese (Reinjerse 2019).

¹⁶ Per una trattazione dei criteri attraverso operare una distinzione tra composti e polirematiche rimandiamo alla lettura di Masini (2019).

di verbo e nome come in *apribottiglie*) o elementi neoclassici (ad es., i composti come *fonologia*), e hanno come tratto distintivo il fatto di non essere saldate insieme da una relazione grammaticale che sia resa esplicita da una marca formale (Bisetto 2004; Bisetto e Scalise 1999; Iacobini 2004; Masini 2019; Masini e Scalise 2012; Scalise 1992). In italiano, la maggior parte dei composti è riconducibile alla classe nominale (ad es., *bassorilievo*, *buttafuori*) e aggettivale (ad es., *agrodolce*, *biancoazzurro*), ma non sono rarissimi i casi di composti verbali (ad es., *manomettere*) e avverbiali (ad es., *malvolentieri*).

Per quel che riguarda la loro classificazione, Bisetto e Scalise (2005) propongono una tripartizione in: 1) composti subordinativi, cioè formati da elementi in una relazione di dipendenza sintattica implicita (es., *capostazione*); 2) composti attributivi, ossia formati da due elementi di cui l'uno svolge la funzione di modificatore dell'altro (es., *pesce spada*); 3) composti coordinativi, vale a dire formati da due elementi legati da una relazione di coordinazione (es., *poeta attore*). Ciascuno di questi tre assi di classificazione può essere ulteriormente articolato sulla base della presenza/assenza dell'elemento testa, ossia l'elemento cardine che trasmette non solo l'assegnazione della categoria grammaticale, ma anche caratterizzazioni di tipo semantico e flessivo. I composti subordinativi, attributivi e coordinativi, infatti, possono essere sia di tipo endocentrico che di tipo esocentrico, laddove con endocentrico intendiamo quei composti con testa interna al costruito (es., *bassorilievo*, *cassaforte*, *divano-letto*) e con esocentrico quei composti che non sono dotati di una testa (es., *portalettere*, *piedi piatti*, *sordomuto*). A ciò aggiungiamo che l'eterogeneità di queste strutture si manifesta anche attraverso la molteplicità di opzioni possibili per la loro resa grafica, potendo presentarsi come univerbate (es., *sottopassaggio*), come saldate per mezzo di un trattino (es., *legge-truffa*), o come due parole separate (es., *camera oscura*).

Come relazionarsi allora con siffatta diversità? La MIPVU propone di valutare il tipo di trattamento da riservare ai composti sulla base di un criterio primariamente fonologico, in accordo al quale l'analizzabilità del costruito come unità lessicale singola o ulteriormente scomponibile nelle sue parti verrebbe a dipendere dalla indicazione che nei dizionari si fornisce circa l'individualità fonologica dei singoli componenti. La MIP, al contrario, propone una soluzione che pare essere semanticamente orientata, suggerendo di trattare come singole unità lessicali quei composti la cui semantica non sia composizionalmente inferibile,

ma di separare i singoli elementi in quei casi in cui il composto risulti essere semanticamente scomponibile. Tuttavia, il criterio di demarcazione proposto dalla MIP non sembra avere una forza pienamente dirimente. Il principio di composizionalità non è scevro da complicazioni ma si scontra, di fatto, con la variabilità semantica dei costrutti stessi, il cui grado di trasparenza può esser soggetto a notevole variazione. Dato questo stato di cose, risulta effettivamente difficile determinare criteri stabili in base ai quali operare la segmentazione testuale dei composti. Nel nostro tentativo di estensione della MIP all'italiano, adottiamo un criterio certamente pragmatico, preferenzialmente lessicografico e che non tenga conto delle diverse rese grafiche con cui un composto può presentarsi. In conclusione, diciamo che: se il composto è riportato a lemma, sarà analizzabile come una singola unità lessicale; se, al contrario, il composto non compare come lemma, gli elementi su cui si regge andranno analizzati separatamente.

4.3 MIP-IT e costruzioni verbali complesse: i verbi sintagmatici

L'espressione *verbi sintagmatici* (Cini 2008; Iacobini & Masini 2006; Simone 1997) nasce come calco dall'inglese *phrasal verbs* (es., *set up, speak up, take over*) e si riferisce a un particolare tipo di struttura polirematica in cui un elemento propriamente verbale (di solito un verbo di movimento generico e ad alta frequenza) si accompagna a una particella dal valore avverbiale variabile (ad es., locativo in *mettere via*, temporale in *fare prima*, di modo in *parlare chiaro*, ecc.). Si tratta di un raggruppamento verbale fortemente variegato al cui interno sono da includere anche verbi sintagmatici seguiti da elementi preposizionali fissi¹⁷ (ad es., *mettere in moto*), verbi pronominali semplici (ad es., *vederci, capirci, starci, esserci e averci*), verbi pronominali multipli (ad es., *prendersela, farsela o intendersela*), verbi sintagmatico-pronominali (ad es., *farsi accanto, farsi avanti, farsi incontro, farsi indietro, farsi sotto, farsela addosso o andarsene via*) e strutture binomiali (ad es., *andare e venire*). Per quel che concerne il grado di coesione tra costituenti, è interessante notare il fatto che i verbi sintagmatici, pur essendo rappresentati graficamente da due parole separate, hanno un comportamento fonologicamente, sintatticamente e semanticamente unitario. Dal punto di vista fonologico, l'elemento verbale e l'elemento avverbiale/preposizionale non sono prosodicamente separa-

¹⁷ Per un approfondimento, si rimanda a Voghera (2004).

bili ma sono uniti in un unico contorno intonazionale (Simone 1997). Dal punto di vista sintattico, la separabilità dei costituenti è possibile in alcuni contesti (ad es., *andare addosso* permette costruzioni come *Gli è andato addosso con la macchina* e *Gli è andato con la macchina addosso*) ma bloccata in altri (ad es., un verbo come *venire meno* permette espressioni come *è venuto meno l'appoggio* ma non **è venuto l'appoggio meno*). Infine, dal punto di vista semantico, il significato complessivo dei verbi sintagmatici non è necessariamente derivabile come somma dei significati dei componenti interni al costrutto (ad es., verbi in cui la particella ha valore direzionale come in *andare fuori*, o verbi in cui la particella ha valore di intensificatore come in *uscire fuori*), ma può assumere sfumature più o meno idiomatiche (ad es., *buttare giù* non significa solo buttare qualcosa dall'alto verso il basso ma anche deprimere o rattristare, e ancora scrivere qualcosa in modo frettoloso). Per quanto riguarda il piano metodologico, il trattamento da riservare ai verbi sintagmatici all'interno di un contesto di annotazione come quello su cui si sorregge la MIP appare una operazione tutt'altro che scontata. Le norme applicative sono state testate sui cosiddetti *phrasal verbs* (ad es., *look up*, *turn on*, ecc.) e suggeriscono di considerare tali strutture come unità lessicali singole e non scomponibili in unità minori (Pragglejaz 2007). Tale scelta è dettata da motivazioni di ordine primariamente semantico, dal momento che non sarebbe possibile segmentare le singole componenti del costrutto sintagmatico senza intaccarne irrimediabilmente il significato complessivo. Diverso atteggiamento sarebbe invece da assumere nei confronti di verbi che siano accompagnati da particelle preposizionali (cdd. *prepositional verbs*) o avverbiali posizionalmente mobili. In questi ultimi due casi, la MIPVU suggerisce di analizzare tali strutture verbali come combinazioni libere formate da unità lessicali indipendenti (Steen et al. 2010). In aggiunta, onde evitare sovrapposizioni indesiderate tra *phrasal verbs* propriamente detti e costruzioni verbali volatili come i *prepositional verbs*, la MIPVU amplia il quantitativo di indicazioni fornendo due criteri di demarcazione: 1) solo i *phrasal verbs* sarebbero esplicitamente codificati nei dizionari¹⁸; 2) nei *phrasal verbs* veri e propri la particella in accompagnamento all'elemento

¹⁸ In realtà, le decisioni prese dai singoli dizionari non sembrano avere un valore necessariamente dirimente. Le diverse risorse lessicografiche sono suscettibili di variabilità di giudizio e potrebbero non sempre essere concordi sulla definizione di *phrasal verb*.

verbale non avrebbe valore preposizionale. Il quadro italiano non sembra essere meno complesso. Come abbiamo illustrato all'inizio del paragrafo, i verbi sintagmatici costituiscono una classe composta, in cui fluttuano verbi sintagmatici con particella avverbiale (ad es., *mettere via*), con particella preposizionale fissa (ad es., *mettere in moto*), pronominali semplici (ad es., *vederci*), multipli (ad es., *farsela*), sintagmatico-pronominali (ad es., *farsi avanti*) e binomiali (ad es., *andare e venire*). Dinanzi a un sistema così eterogeneo, e in assenza di una lista di riferimento stabile contenente tutte le possibili strutture sintagmatiche disponibili in italiano (Masini et al. 2020), un criterio affidabile potrebbe essere quello lessicografico-semantic. Con ciò intendiamo dire che, al pari di quanto stabilito per il trattamento delle altre strutture polirematiche (par. 4.1.), proponiamo di considerare come singole unità lessicali quelle combinazioni verbali stabili che siano state riportate a lemma e che mostrino di avere un sovrappiù semantico rispetto ai significati dei singoli costituenti (ad es., *dare addosso a qualcuno* nel senso di *inferire contro qualcuno*), ma consideriamo come segmentabili in componenti minori quelle espressioni per cui, al contrario, non sia prevista una lemmatizzazione e il significato complessivo dell'espressione sia ricostruibile composizionalmente (ad es., *uscire fuori*).

4.4 MIP-IT e costruzioni verbali complesse: i verbi pronominali

I verbi pronominali (Cennamo 1993; Lo Cascio 1970; Jezek 2004, 2011; Masini 2012; Serianni 1988) sono costruzioni verbali caratterizzate dalla presenza di un pronome clitico (o più di uno come in *andarsene*), che può presentarsi come intrinsecamente saldato al lemma (ad es., *pentirsi*) o opzionalmente combinato con la forma base de verbo (ad es., *riempire vs. riempirsi*). All'interno di questa ampia e variegata classe di costruzioni, il clitico *-sì* è quello che sembra mostrare la maggiore polifunzionalità (Masini 2012), potendo figurare in strutture riflessive (ad es., *asciugarsi*), riflessive indirette (ad es., *tagliarsi*), intensive o espressive (ad es., *mangiarsi*), reciproche (ad es., *salutarsi*), reciproche indirette (ad es., *scambiarsi la borsa*) e intransitive pronominali (ad es., *ingelosirsi*). La macro-classe dei verbi pronominali include anche i cosiddetti verbi procomplementari (Jezek 2011), ovvero costruzioni che nascono dalla combinazione tra un lemma verbale e particelle clitiche come *-ci* (ad es., *arrivarci*), *la* (ad es., *finirla*), *le* (ad es., *suonarle*), *-ne* (ad es., *volerne*), o dalla combinazione con due

o più clitici (ad es., *farcela*). È interessante notare che il significato dei verbi procomplementari si discosta da quello del lemma verbale base (cioè privo di particella clitica). Ciò significa che in un verbo procomplementare come *casarci* è la particella stessa a determinare in maniera inequivocabile il senso di “cadere in un tranello o in uno scherzo, farsi abbindolare” e a separare semanticamente questo verbo dalla forma base *casare*, la quale indica il più generico “cadere lungo un percorso”. Tenuto conto di una così abbonante varietà verbale, come adattare le linee guida MIP (e integrazioni fornite dalla MIPVU) all’italiano, assicurando una corretta segmentazione dei verbi pronominali? In questo contesto, può esser utile modellare la nostra proposta a partire da riflessioni calibrate su una lingua come il francese che sembra presentare problemi decisamente assimilabili a quelli proposti per l’italiano. In particolare, Reijnjerse (2019) consiglia di considerare come due unità lessicali indipendenti il lemma verbale e l’elemento clitico, laddove quest’ultimo abbia un valore riflessivo (ad es., *laver* “lavare” vs. *se laver* “lavarsi”) o reciproco (*ad es., voir* “vedere” vs. *se voir* “vedersi”) e dunque non comporti una modificazione semantica del verbo. Nello stesso studio, Reijnjerse (2019) aggiunge che, in tutti quei casi in cui il clitico francese *se* non assume né valore riflessivo né valore reciproco, verbo e particella dovrebbero essere tenuti insieme e considerati come un’unica unità lessicale. Tale proposta sarebbe corroborata dalla presenza di verbi come *se méfier* “insospettirsi”, in cui il pronome non è né opzionale né separabile dall’elemento verbale. Stesso trattamento (singola unità lessicale) sarebbe da riservare a tutti quei casi in cui tra la forma base di un verbo (ad es., *battre* “battere, colpire”) e la sua alternativa pronominale (ad es., *se battre* “battersi”) non ci sia equivalenza semantica, né sul piano concreto né sul piano metaforico. Per quel che concerne l’italiano, riteniamo che possa essere utile proporre un approccio tripartito simile a quello adottato per il francese, in virtù del quale sarà necessario operare delle discriminazioni di tipo pratico: 1) in tutti quei casi in cui il clitico non alteri la semantica del verbo, si preferirà separare gli elementi interni alla costruzione verbale (ad es., *lavar-si*); 2) in tutti quei casi in cui la presenza del clitico sia obbligatoria, si preferirà tenere insieme ambo gli elementi della costruzione (ad es., *insospettirsi* e non *insospettir-si*); 3) in tutti quei casi in cui la semantica del verbo pronominale non sia equipollente rispetto alla semantica del verbo base, si preferirà considerare la costruzione verbale

pronominale come un'unica unità lessicale (ad es., *battersi* e non *batter-si*). Le differenziazioni procedurali qui proposte riguardano tutte quelle costruzioni pronominali in cui la particella d'accompagnamento al verbo sia presente in posizione enclitica. Nel caso in cui la particella si presentasse in posizione proclitica, andando dunque a precedere il lemma verbale, riteniamo che la soluzione migliore possa essere quella di tenere separate le singole unità e analizzare gli elementi nella loro discontinuità sintattica (ad es., */farcela/ ma / ce / la / fai*).

5. Dalla MIP alla MIP-IT

Sulla base delle peculiarità grammaticali, sintattiche e semantiche discusse nel paragrafo precedente, presentiamo di seguito una preliminare proposta di adattamento della MIP/MIPVU all'italiano (MIP-IT). La MIP-IT si snoda lungo lo stesso percorso procedurale proposto per la MIP (Pragglejaz 2007), ma aggiunge indicazioni più specifiche per la corretta segmentazione del testo in unità lessicali, step preliminare per l'applicazione dell'iter di analisi. Adottiamo come regola generale quella di segmentare il testo in unità lessicali singole, considerando come tali tutti quegli elementi che sono dotati di specifica appartenenza grammaticale (PoS) e che compaiano come entrate o lemmi all'interno del dizionario. Le eccezioni alla regola generale sono rappresentate da:

- 1) Polirematiche: se la polirematica compare come entrata a parte nel dizionario, sarà da analizzare come singola unità lessicale.
- 2) Composti: se il composto compare come lemma nel dizionario, sarà da analizzare come singola unità lessicale.
- 3) Verbi sintagmatici: i verbi sintagmatici che compaiono a lemma all'interno del dizionario saranno da analizzare come costrutti coesi e non ulteriormente scomponibili.
- 4) Verbi pronominali: sono da considerare come strutture unitarie e non separabili quelle strutture verbali pronominali a) che compaiono come lemmi nel dizionario e b) in cui il clitico sia obbligatorio e non separabile dalla forma verbale o c) in cui l'aggiunta del clitico apporti una modificazione semantica sostanziale.

Per quel che concerne la scelta dei riferimenti lessicografici adottati in contesto italiano, come già precedentemente specificato, ci avvarremo delle ver-

sioni online del GRADIT¹⁹ (De Mauro 1999) e, come fonte supplementare, del *Vocabolario Treccani*²⁰.

5.1 MIP-IT: Discussione di un caso di studio

Al fine di illustrare come poter applicare praticamente la MIP all'italiano (MIP-IT), proponiamo un esempio di analisi su una porzione di testo estratto da un mini-corpus contenente trascrizioni di parlato spontaneo di soggetti italo-foni che hanno sofferto di anoressia nervosa. I trascritti sono stati automaticamente ricavati a partire da video liberamente accessibili sulla piattaforma TikTok (Donati et al. 2023). Il dizionario che abbiamo utilizzato per approntare l'analisi è il GRADIT (versione online). Tuttavia, durante l'esercizio di annotazione abbiamo consultato il *Vocabolario Treccani* come fonte supplementare:

(4) a volte quando quella maledetta voce che ha trovato dolce dimora nella mia testa smette di blaterare io amo affacciarmi alla finestra osservare il mondo che mi circonda.

a volte / quando / quella / maledetta / voce / che / ha / trovato / dolce / dimora / nella / mia / testa / smette / di / blaterare / io / amo / affacciar (e) / -mi / alla / finestra / osservare / il / mondo / che / mi / circonda/.

Di seguito, lo svolgimento dell'analisi (con necessarie modifiche ricalibrate sull'italiano):

A volte

Significato contestuale: locuzione avverbiale col significato di “di tanto in tanto, talvolta”. In questo contesto, la locuzione sta ad indicare che la situazione di cui parla non si verifica sempre ma solo in alcune circostanze.

Significato base: non c'è un significato più basilare.

Significato contestuale vs. significato base: Il significato contestuale e il significato base coincidono.

Metaforico? No.

¹⁹ <https://dizionario.internazionale.it>.

²⁰ <https://www.treccani.it/vocabolario/>.

Quando

Significato contestuale: congiunzione con valore temporale che qui serve a introdurre una proposizione secondaria di valore, per l'appunto, temporale.

Significato base: non c'è un significato più basilare.

Significato contestuale vs. significato base: Il significato contestuale e il significato base coincidono.

Metaforico? No.

Quella

Significato contestuale: aggettivo dimostrativo che indica una cosa nominata precedentemente o già nota a chi ascolta. In questo contesto, si riferisce alla voce della malattia che prende il controllo dei pensieri del soggetto narrante.

Significato base: non c'è un significato più basilare.

Significato contestuale vs. significato base: Il significato contestuale e il significato base coincidono.

Metaforico? No.

Maledetta

Significato contestuale: aggettivo che, per estensione, si riferisce a qualcosa che è causa anche indirettamente di contrarietà o rabbia. In questo contesto, la malattia è maledetta in quanto fonte di stati emotivi negativi.

Significato base: aggettivo con valore di “segnato dalla maledizione divina, legato a un destino di perdizione conseguente alla maledizione divina”.

Significato contestuale vs. significato base²¹: il significato contestuale diverge dal significato base ma non può essere compreso per via comparativa.

Metaforico? No.

²¹ In realtà, questo potrebbe essere considerato un caso limite tra la metafora (A è B) e la metonimia (A sta per B). Non escludiamo, infatti, che il meccanismo alla base dello slittamento di significato possa fondarsi su un processo sostitutivo, più che comparativo, in virtù del quale il soggetto/oggetto (la malattia) che subisce una certa azione (in questo caso la maledizione) sta per l'entità che è causa dello stato emotivo negativo (rabbia) connesso all'azione stessa (l'azione del maledire). Per un approfondimento, si rimanda a Goosens (1990) e Panther & Radden (1999).

Voce

Significato contestuale: in questo contesto, indica l'impulso interiore (si tratta dell'anoressia nervosa) che spinge a comportarsi in un certo modo e ad adottare tutta una serie di comportamenti alimentari (ma non solo) malsani.

Significato base: indica il "suono emesso dagli esseri umani o da altri animali per produrre segnali, cantare, parlare, sfruttando il passaggio dell'aria attraverso la gola e la bocca o strutture analoghe nei volatili".

Significato contestuale vs. significato base: il significato contestuale diverge dal significato base e può essere compreso per via comparativa. Comprendiamo la malattia nei termini di una entità animata (processo di personificazione)²² e i pensieri perturbanti causati dalla malattia nei termini di flussi sonori che risuonano nella testa.

Metaforico? Sì.

*Che*²³

Significato contestuale: pronomi relativo utilizzato per riferirsi a persona o cosa, generalmente usato con valore di soggetto o complemento oggetto. In questo caso si riferisce alla voce della malattia.

Significato base: non c'è un significato più basilare.

Significato contestuale vs. significato base: Il significato contestuale e il significato base coincidono.

Metaforico? No.

*Ha*²⁴

Significato contestuale: in questo contesto, svolge funzione di ausiliare nei tempi composti (passato prossimo) di un verbo transitivo attivo.

²² La MIPVU propone di adottare un tag specifico per questo tipo di casi, ossia PP (eng. *possible personification*).

²³ Se avessimo qui seguito pedissequamente il protocollo MIPVU (e non MIP), avremmo dovuto considerare come metaforico anche l'uso contestuale del pronome relativo *che*. Tale elemento, infatti, avrebbe attivato una metafora implicita. Tuttavia, come già specificato nel testo, qui prendiamo in considerazione le sole metafore indirette.

²⁴ Ricordiamo che la MIP tratta gli ausiliari come elementi dal valore semantico schematico e astratto. Ciò significa che, all'interno di questo protocollo, tali elementi non sono mai considerati come portatori di significato metaforico. Per un approfondimento si rimanda alla lettura di Pragglejaz (2007).

Significato base: non c'è un significato più basilare.

Significato contestuale vs. significato base: Il significato contestuale e il significato base coincidono.

Metaforico? No.

Trovato

Significato contestuale: In questo caso, il verbo trovare ha valore di “procurarsi, riuscire ad avere”.

Significato base: Il GRADIT indica come significato base quello di “rinvenire una cosa che si cerca, che si desidera, di cui si vuole disporre”.

Significato contestuale vs. significato base²⁵: Pur essendoci una lieve sfumatura tra i due sensi non ci sembra che ci sia una divergenza sostanziale tra i due significati e, dunque, non ci sembra che tale differenza possa essere intesa per via comparativa.

Metaforico? No.

Dolce

Significato contestuale: in questo contesto l'aggettivo assume un valore astratto, figurato, e sta a esplicitare una sensazione o un atteggiamento di affetto verso un qualcosa che il soggetto avverte come a lei caro.

Significato base: Il GRADIT ci ricorda che l'aggettivo dolce è primariamente utilizzato in tutti quei casi in cui ci si riferisce a qualcosa che “ha il sapore gradevole proprio dello zucchero o del miele in contrapposizione all'amaro”.

Significato contestuale vs. significato base: il significato contestuale diverge dal significato base e può essere compreso per via comparativa. In questo caso intendiamo la sensazione di piacevolezza legata all'idea di dimora stabile in cui l'anoressia possa albergare nei termini di una percezione sensoriale specifica, associata all'apparato gustativo.

Metaforico? Sì

²⁵ Casi dubbi come questo (o come il caso precedentemente discusso in cui ci sembrava esserci una sorta di interferenza tra metafora e metonimia) andrebbero seriamente risolti mediante il confronto con altri annotatori. Ricordiamo infatti che uno dei punti di forza della MIP consiste nello sfruttamento di strategie attraverso cui misurare l'accordo tra annotatori, sia da un punto di vista quantitativo che qualitativo (Pragglejaz 2007).

Dimora

Significato contestuale: In questo contesto non si tratta di un luogo reale e fisico ma di uno spazio figurato.

Significato base: indica il luogo dove si abita.

Significato contestuale vs. significato base: il significato contestuale diverge dal significato base e può essere compreso per via comparativa. In questo caso, la testa rappresenta una sorta di spazio abitativo astratto dove trovano dimora i pensieri perturbanti causati dalla malattia da cui è affetto il soggetto narrante.

Metaforico? Sì

Nella

Significato contestuale: indica una relazione spaziale dal valore figurato, dal momento che la voce di cui parla il soggetto non è una presenza reale e fisicamente collocabile all'interno di un luogo concreto e spazialmente definito.

Significato base: si tratta di una preposizione articolata formata da *in* e *il* che esplicita la relazione tra due oggetti di cui uno è contenuto nell'altro.

Significato contestuale vs. significato base: il significato contestuale diverge dal significato base e può essere compreso per via comparativa. In questo caso, tale preposizione serve a introdurre una concettualizzazione su base spaziale, in virtù della quale la testa è rappresentata come una sorta di contenitore per pensieri ed emozioni negativamente connotati ma associabili alla malattia.

Metaforico? Sì

Mia

Significato contestuale: aggettivo possessivo che sta a indicare una relazione di proprietà, possesso, appartenenza all'individuo in sé.

Significato base: non c'è un significato più basilare.

Significato contestuale vs. significato base: Il significato contestuale e il significato base coincidono.

Metaforico? No

Testa

Significato contestuale: il soggetto narrante si riferisce propriamente alla parte superiore o anteriore del corpo che è anche sede degli organi preposti alle funzioni intellettive, ma considera tale parte del corpo alla stregua di un contenitore al cui interno alberga la maledetta voce della malattia.

Significato base: parte superiore o anteriore del corpo che è anche sede degli organi preposti alle funzioni intellettive.

Significato contestuale vs. significato base: il significato contestuale diverge dal significato base e può essere compreso per via comparativa. Intendiamo la testa nei termini di un contenitore concreto entro cui collocare spazialmente entità astratte che prendono la forma di pensieri ossessivi ricorrenti.

Metaforico? Sì

Smette

Significato contestuale: sospendere, cessare temporaneamente o definitivamente di fare o dire qualcosa.

Significato base: non c'è un significato più basilare.

Significato contestuale vs. significato base: Il significato contestuale e il significato base coincidono.

Metaforico? No

Di²⁶

Significato contestuale: preposizione che, in questo caso specifico, serve da collegamento tra due elementi verbali.

Significato base: non c'è un significato più basilare.

²⁶ La MIP sottolinea la difficoltà nel trovare uno scarto tra significato base e significato contestuale di parole grammaticali come le preposizioni (Pragglejaz 2007). Il motivo, proprio come nel caso precedentemente discusso degli ausiliari, è che tali elementi hanno significati tendenzialmente astratti. Ad eccezione di qualche raro caso (ad es., una preposizione come *in* indica primariamente relazioni di tipo spaziale), il protocollo non associa un valore metaforico a questi elementi grammaticali.

Significato contestuale vs. significato base: Il significato contestuale e il significato base coincidono.

Metaforico? No

Blaterare

Significato contestuale: In questo caso non ci si riferisce a un vero e proprio flusso sonoro quanto piuttosto a un flusso di pensieri pervasivi e perturbanti.

Significato base: questo verbo indica un particolare modo di parlare: rumoroso, continuo e fastidioso.

Significato contestuale vs. significato base: il significato contestuale diverge dal significato base e può essere compreso per via comparativa. In questo caso specifico, infatti, intendiamo la forza e l'intensità dei sintomi (che qui si presentano come pensieri negativi costanti) nei termini di una forza sonora e vocalica continua, assordante.

Metaforico? Sì

Io

Significato contestuale: il pronome soggetto indica la persona che parla quando si riferisce a sé stessa. In questo caso specifico, il soggetto affetto da anoressia nervosa racconta la propria esperienza di vita e alimentare.

Significato base: non c'è un significato più basilare.

Significato contestuale vs. significato base: Il significato contestuale e il significato base coincidono.

Metaforico? No

Amo

Significato contestuale: in questo caso ci sembra che il significato contestuale corrisponda alla terza accezione elencata nel GRADIT. Il verbo significherebbe “prediligere, provare interesse”.

Significato base: Sebbene come prima accezione sia segnalata quella di “provare amore per qualcuno, voler bene”, non ci sembra che tra le due diverse modulazioni del significato del lemma ci sia una differenza tale da poter dire che un significato sia effettivamente più basilare dell'altro.

Significato contestuale vs. significato base: Il significato contestuale e il significato base coincidono.

Metaforico? No

Affacciare

Significato contestuale: il verbo ha un valore azionale e sta ad indicare lo sporgersi o mostrarsi da una finestra, una porta e simili.

Significato base: non c'è un significato più basilare.

Significato contestuale vs. significato base: Il significato contestuale e il significato base coincidono.

Metaforico? No

-mi

Significato contestuale: forma atona di io, come complemento oggetto, qui in posizione enclitica, di valore riflessivo.

Significato base: non c'è un significato più basilare.

Significato contestuale vs. significato base: Il significato contestuale e il significato base coincidono.

Metaforico? No

Alla

Significato contestuale: preposizione articolata formata da più il indicante moto.

Significato base: non c'è un significato più basilare.

Significato contestuale vs. significato base: Il significato contestuale e il significato base coincidono.

Metaforico? No

Finestra

Significato contestuale: corrisponde effettivamente ad una apertura nel muro di un edificio destinata a far prendere aria e luce agli ambienti interni.

Significato base: non c'è un significato più basilare.

Significato contestuale vs. significato base: Il significato contestuale e il significato base coincidono.

Metaforico? No

Osservare

Significato contestuale: il verbo qui sta a significare proprio il guardare con attenzione qualcosa.

Significato base: non c'è un significato più basilare.

Significato contestuale vs. significato base: Il significato contestuale e il significato base coincidono.

Metaforico? No

*Il*²⁷

Significato contestuale: articolo determinativo con cui si individua un particolare componente di una classe.

Significato base: non c'è un significato più basilare.

Significato contestuale vs. significato base: Il significato contestuale e il significato base coincidono.

Metaforico? No

Mondo

Significato contestuale: indica la totalità delle cose esistenti in natura e che, in questo contesto specifico, rientrano all'interno del raggio visivo del soggetto narrante.

Significato base: non c'è un significato più basilare.

²⁷ All'interno del protocollo MIP, gli articoli, così come gli ausiliari e molte congiunzioni, non assumono mai velature metaforiche.

Significato contestuale vs. significato base: Il significato contestuale e il significato base coincidono.

Metaforico? No

Che

Significato contestuale: pronome relativo *che*, in questo contesto specifico, si riferisce al mondo, qui usato in funzione di soggetto.

Significato base: non c'è un significato più basilare.

Significato contestuale vs. significato base: Il significato contestuale e il significato base coincidono.

Metaforico? No

Mi

Significato contestuale: forma atona di *io*, come complemento oggetto, qui in posizione proclitica.

Significato base: non c'è un significato più basilare.

Significato contestuale vs. significato base: Il significato contestuale e il significato base coincidono.

Metaforico? No

Circonda

Significato contestuale: indica proprio il limitare tutt'intorno, il cingere uno spazio.

Significato base: non c'è un significato più basilare.

Significato contestuale vs. significato base²⁸: Il significato contestuale e il significato base coincidono.

Metaforico? No

²⁸ Tuttavia, si potrebbe obiettare che casi simili possano essere fatti rientrare più propriamente nella classe dei cosiddetti movimenti fittivi (eng. Fictive motion). Per un approfondimento, si rimanda a Talmy (1996) e Matlock (2004).

6. Conclusioni

Il presente lavoro si pone come un preliminare tentativo di esplorazione delle possibilità e delle criticità poste dal protocollo MIP in un campo di applicazione nuovo, ossia l'italiano, in cui tale modello non ha ancora trovato adozione (o forse fortuna). Gli obiettivi della ricerca possono essere ulteriormente esplicitati su un duplice livello. Da un lato, abbiamo provato a discutere le particolarità grammaticali e linguistiche proprie dell'italiano che necessitano di una maggiore attenzione al fine di poter perfezionare un modello principalmente collaudato sull'inglese e che richiede l'implementazione di una serie di modifiche (anche se non troppo sostanziali) per poter diventare uno strumento di annotazione metaforica valido anche per l'annotazione di dati italiani. Dall'altro lato, abbiamo provato a fornire una dimostrazione pratica, sulla base di un esempio di dati reali estratti da corpus, su come poter effettivamente sfruttare questo stesso protocollo al fine di produrre una analisi metaforica affidabile e, soprattutto, replicabile in altri e successivi studi sull'italiano. Difatti, l'analisi qui proposta, snodandosi sia sul piano della discussione teorica sia sul piano concreto dell'uso pratico, non solo testimonia l'importanza di linee guida specifiche che regolino in modo esplicito e coerente l'applicazione del protocollo MIP all'italiano, ma rappresenta anche una sorta di punto zero, certamente suscettibile di critica e revisione, per eventuali futuri tentativi di applicazione del medesimo modello. Tale estensione potrebbe rivelarsi utile per ricercatori interessati allo studio della metafora in campi di ricerca anche molto diversi tra loro (analisi del discorso ma anche contesti sperimentali come quello proprio della psicolinguistica), fornendo non solo degli strumenti di analisi affidabili ma permettendo anche una replicabilità e una comparabilità (anche in ottica quantitativa) dei risultati ottenuti.

Riferimenti bibliografici

- Aristotele, *Poetica*, trad. it. a cura di Lanza, D., Milano, BUR (2011).
Aristotele, *Retorica*, trad. it. a cura di Dorati, M., Milano, Mondadori (2009).
Beardsley, Monroe, 1962, "The metaphorical twist", *Philosophy and Phenomenological Research* 22 (3), 293-307.

- Bisetto, Antonietta, 2004, "Composizione con elementi italiani", In Grossmann M. & Rainer F. (eds.), *La formazione delle parole in italiano*, Tübingen, Niemeyer, 33-51.
- Bisetto, Antonietta & Scalise, Sergio, 1999, "Compounding. Morphology and/or syntax?", In Mereu, L. (ed.), *Boundaries of Morphology and Syntax*, Amsterdam/Philadelphia, Benjamins, 31-48.
- Bisetto, Antonietta & Scalise, Sergio, 2005, "Classification of Compounds", *Lingue e Linguaggio* 2, 319-332.
- Black, Max, 1954, "Metaphor", *Proceedings of the Aristotelian society* 55, 273-294.
- Black, Max, 1962, *Models and metaphors*, Ithaca, NY, Cornell University Press.
- Black, Max, 1979/1993, "More about metaphor", In Ortony A. (ed.), *Metaphor and thought: Second edition*, Cambridge, Cambridge University Press, 19-43.
- Cameron, Lynne, 1999, "Operationalising 'metaphor' for applied linguistic research", In Cameron L. & Low G. (eds.), *Researching and applying metaphor*, Cambridge, Cambridge University Press, 3-28.
- Cameron, Lynne, 2003, *Metaphor in educational discourse*, London and New York, Continuum.
- Cameron, Lynne & Maslen, Robert, 2010, "Identifying metaphors in discourse data", In Cameron L. & Maslen R. (eds.), *Metaphor analysis: Research practice in applied linguistics, social sciences and the humanities*, London, Equinox, 97-115.
- Cennamo, Michela, 1993, *The Reanalysis of Reflexives: a Diachronic Perspective*, Napoli, Liguori Editore.
- Charteris-Black, Jonathan, 2000, "Metaphor and vocabulary teaching in ESP economics", *English for Specific Purposes* 19, 149-165.
- Charteris-Black, Jonathan, 2004, *Corpus approaches to critical metaphor analysis*, London, Palgrave MacMillan.
- Cienki, Alan, 1998, "Metaphoric gestures and some of their relations to verbal metaphoric expressions", In Koenig J.-P. (ed.), *Discourse and Cognition: Bridging the Gap*, Stanford, CA, Center for the Study of Language and Information, 189-204.
- Cienki, Alan, 2008, "The application of conceptual metaphor theory to political discourse: Methodological questions and some possible solutions", In Carver T. & Pikalo J. (eds.), *Political language and metaphor: Interpreting and changing the world*, London, Routledge, 241-256.
- Cienki, Alan, 2017, "Analysing metaphor in gesture: A set of metaphor identification guidelines for gesture (MIG-G)", In Semino E. & Demjén Z. (eds.), *The Routledge handbook of metaphor and language*, London, Routledge, 131-147.
- Cienki, Alan & Müller, Cornelia, 2008, *Metaphor and Gesture*, Amsterdam, John Benjamins Publishing Company.

- Cini, Monica, 2008, I verbi sintagmatici in italiano: una proposta di descrizione, In Cini M. (ed), *Atti delle giornate di studio* (Torino, 19-20 febbraio 2007), Frankfurt am Main, Lang, 41-59.
- De Mauro, Tullio, 1999, "Introduzione", GRADIT 1999-2007, vol. 1^o, VII-XLII.
- De Mauro, Tullio & Voghera, Miriam, 1996, "Scala mobile. Un punto di vista sui lessemi complessi", In Benincà P. (ed), *Italiano e dialetti nel tempo. Saggi di grammatica per Giulio C. Lepschy*, Roma, Bulzoni, 99-131.
- Donati, Melissa, Vernillo, Paola, Polidori, Ludovica, Gagliardi Gloria, 2023, *RAC - Recovery from Ana/Anorexia Corpus*, ILC-CNR for CLARIN-IT repository hosted at Institute for Computational Linguistics "A. Zampolli", National Research Council, Pisa".
- Geeraerts, Dirk, 2010, *Theories of lexical semantics*, Oxford, Oxford University Press.
- Gibbs, R. W., Jr., 1994, *The poetics of mind: Figurative thought, language, and understanding*, Cambridge, Cambridge University Press.
- Gibbs, R.W., Jr., 2008, *The Cambridge handbook of metaphor and thought*. New York, Cambridge University Press.
- Gibbs, R. W., Jr. & Colston, H. L., 2012, *Interpreting figurative meaning*, Cambridge, Cambridge University Press.
- Grady, Joe, 1997, *Foundations of meaning: Primary metaphors and primary scenes*, unpublished thesis, University of California, Berkeley, CA.
- Goossens, Louis, 1990, "Metaphonymy: the interaction of metaphor and metonymy in expressions for linguistic action", *Cognitive Linguistics* 1 (3), 323-340.
- Iacobini, Claudio, 2004, "Composizione con elementi neoclassici", In Grossmann M. & Rainer F. (eds), *La formazione delle parole in italiano*, Tübingen, Niemeyer, 69-95.
- Iacobini, Claudio & Masini, Francesca, 2006, "The emergence of verb-particle constructions in Italian: locative and actional meanings", *Morphology* 16 (2), 155-188.
- Jezek, Elisabetta, 2004, "Interazioni tra aspetto e diatesi nei verbi pronominali italiani", *Studi di grammatica italiana* 22, 239-281.
- Kövecses, Zoltan, 2011, "Methodological issues in conceptual metaphor theory", In Handl S. & Schmid H.-J. (eds.), *Windows to the mind: Metaphor, metonymy and conceptual blending*, Berlin, Mouton de Gruyter, 23-40. <https://doi.org/10.1515/9783110238198.23>
- Lakoff, George, 1990, "The invariance hypothesis: Is abstract reason based on image-schemas?", *Cognitive Linguistics* 1, 39-71.
- Lakoff, George, 1993, "The contemporary theory of metaphor", In Ortony A. (ed), *Metaphor and thought: Second edition*, Cambridge, Cambridge University Press, 202-251.
- Lakoff, George & Johnson, Mark, 1980, *Metaphors we live by*. Chicago, Chicago University Press.

- Levin, Samuel, 1977, *The semantics of metaphor*, Baltimore, Md, Johns Hopkins University Press.
- Little, William, Fowler, Henry Watson, Coulson, Jessie & Onions, Charles Talbut (Eds.), 1973, *The shorter Oxford dictionary on historical principles (3rd ed.)*, Oxford, UK, Clarendon.
- Littlemore, Jeanette, Krennmayr, Tina, Turner, James & Turner, Sarah, 2013, "An investigation into metaphor use at different levels of second language writing", *Applied Linguistics* 35(2), 117-144. <https://doi.org/10.1093/applin/amt004>
- Lo Cascio, Vincenzo, 1970, *Strutture pronominali e verbali italiane*, Bologna, Zanichelli.
- Loewenberg, Ina, 1975, "Identifying metaphors", *Foundations of Language* 12, 315-38.
- Masini, Francesca, 2009, "Phrasal lexemes, composunds and phrases", *Word Structure* 2(2), 254-71.
- Masini, Francesca, 2011, "Polirematiche, parole", *Enciclopedia dell'Italiano, 2: M-Z, ROMA, Istituto dell'Enciclopedia Italiana G. Treccani*, 1109-1112 [voce di enciclopedia/dizionario].
- Masini, Francesca, 2012, *Parole sintagmatiche in italiano*, Roma, Caissa Italia.
- Masini, Francesca, 2019, "Compounds and multi-word expressions in Italian", In Schlücker, B. (ed), *Complex Lexical Units. Compounds and Multi-Word Expressions*, Berlino, De Gruyter, 153-187.
- Masini, Francesca & Scalise, Sergio, 2012, "Italian compounds", *Probus* 24 (1), 61-91.
- Masini, Francesca, Micheli, Silvia, Zaninello, Andrea, Castagnoli, Sara & Nissim, Malvina, 2020, "Multiword expressions we live by: a validated usage-based dataset from corpora of written Italian", *Proceedings of the Seventh Italian Conference on Computational Linguistics, Aachen, CEUR Workshop Proceedings, «CEUR WORKSHOP PROCEEDINGS»*, (atti di: CLiC-it 2020 – Italian Conference on Computational Linguistics 2020, Bologna (online), March 1-3, 2021), 1-5.
- Matlock, Teenie, 2004, "The conceptual motivation of fictive motion", In Radden G. & Panther K. (eds), *Studies in Linguistic Motivation*, New York/Berlin, Mouton de Gruyter, 221-248.
- McNeill, David, 1992, *Hand and mind*. Chicago, University of Chicago Press.
- Nacey, Susan 2013, *Metaphors in learner English*, Amsterdam, John Benjamins. <https://doi.org/10.1075/milcc.2>
- Nacey, Susan, Dorst, Aletta, Krennmayr, Tina & Reijnierse Gudrun (Eds.), 2019, *Metaphor identification in multiple languages: MIPVU around the world*, Amsterdam, John Benjamins.
- Ortony, Anthony, (Ed.), 1979/1993, *Metaphor and thought: Second edition*, Cambridge, Cambridge University Press.

- Panther, Klaus-Uwe & Radden, Günter, 1999, *Metonymy in Language and Thought*, Amsterdam, John Benjamins.
- Pragglejaz Group, 2007, "MIP: A method for identifying metaphorically used words in discourse", *Metaphor and Symbol* 22, 1-39.
- Reijnierse, Gudrun, 2019, "Linguistic metaphor identification in French", In Nacey S., Dorst A., Krennmayr, T. & Reijnierse G. (eds), *Metaphor identification in multiple languages: MIPVU around the world*, Amsterdam, John Benjamins, 69-90.
- Ricoeur, Paul, [1975] 2010, *La métaphore vive*, Paris, Seuil [Tr. It. *La metafora viva. Dalla retorica alla poetica: per un linguaggio di rivelazione*, Milano, Jaca Book].
- Rundell, Michale, & Fox, Gwyneth (Eds.), 2002, *Macmillan English dictionary for advanced learners*, Oxford, UK, Macmillan Education.
- Serianni, Luca, 1988, *Grammatica italiana. Italiano comune e lingua letteraria. Suoni, forme, costrutti*, con la collaborazione di Alberto Castelvechi, Torino, UTET.
- Scalise, Sergio, 1992, "Compounding in Italian", *Italian Journal of Linguistics/Rivista di Linguistica* 4 (1), 175-199.
- Schmitt, Rudolph, 2005, "Systematic metaphor analysis as a method of qualitative research", *The Qualitative Report* 10, 358-394.
- Searle, John, 1979/1993, "Metaphor", In Ortony A. (ed), *Metaphor and thought: Second edition*, Cambridge, Cambridge University Press, 92-123.
- Semino, Elena, 2008, *Metaphor in discourse*, Cambridge, Cambridge University Press.
- Semino, Elena, 2019, "Afterword: Some reflections on MIPVU across languages", In Nacey S., Dorst A., Krennmayr, T. & Reijnierse G. (eds), *Metaphor identification in multiple languages: MIPVU around the world*, Amsterdam, John Benjamins, 3131-321.
- Semino, Elena, Heywood, John & Short, Mick, 2004, "Methodological problems in the analysis of metaphors in a corpus of conversations about cancer", *Journal of Pragmatics* 36, 1271-1294. <https://doi.org/10.1016/j.pragma.2003.10.013>
- Simone, Raffaele, 1997, "Esistono verbi sintagmatici in italiano?", In De Mauro T. & Lo Cascio V. (ed), *Lessico e grammatica. Teorie linguistiche e applicazioni lessicografiche, Atti del Congresso interannuale della Società di Linguistica Italiana* (Madrid, 21-25 febbraio 1995), Roma, Bulzoni, 155-170.
- Simone, Raffaele, 2006, "Classi di costruzioni", in Grandi N. & Iannàccaro G. (ed), *Scritti in onore di Emanuele Banfi in occasione del suo 60° compleanno*, Cesena/Roma, Caissa, 383-409.
- Simone, Raffaele, 2008, "I verbi sintagmatici come costruzione e come categoria", In Cini M. (ed), *Atti delle giornate di studio* (Torino, 19-20 febbraio 2007), Frankfurt am Main, Lang, 13-30.
- Simone, Raffaele & Piunno, Valentina, 2017, "Entry word combination: lexicographical representation and lexicological aspects", *Studi e Saggi Linguistici* 55(2), 13-44.

- Steen, Gerard, 1999, "Metaphor and discourse: Towards a linguistic checklist for metaphor analysis", In Cameron L. & Low G. (eds), *Researching and applying metaphor*, Cambridge, Cambridge University Press, 81-104.
- Steen, Gerard, 2002, "Metaphor identification: A cognitive approach", *Style* 36, 386-407.
- Steen, Gerard, 2002, "Towards a procedure for metaphor identification", *Language and Literature* 11, 17-33.
- Steen, Gerard, 2008, "The paradox of metaphor: Why we need a three-dimensional model of metaphor", *Metaphor & Symbol* 23 (4), 213-241.
- Steen, Gerard, 2007, *Finding metaphor in grammar and usage*, Amsterdam, John Benjamins. <https://doi.org/10.1075/celcr.10>
- Steen, Gerard, 2017, "Identifying metaphors in language", In Semino E. & Demjén Z. (eds), *The Routledge handbook of metaphor and language*, London, Routledge 73-87.
- Steen, Gerard, Dorst, Aletta, Herrmann, Berenike, Kaal, Anna, Krennmayr, Tina & Pasma, Trijntje, 2010, *A method for linguistic metaphor identification: From MIP to MIPVU*, Amsterdam, John Benjamins. <https://doi.org/10.1075/celcr.14>
- Talmy, Leonard, 1996, "Fictive motion in language and 'ception'", In Bloom P., Peterson M., Nadel A. L. & Garrett M. F. (eds), *Language and Space*, Cambridge, MA, The MIT Press, 211-276.
- Voghera, Miriam, 2004, "Polirematiche", in Grossmann M. & Rainer F. (eds), *La formazione delle parole in italiano*, Tübingen, Niemeyer, 56-69.

BIBLIOGRAFIE, RECENSIONI, RASSEGNE

**Giuliana Giusti, Vincenzo Nicolò Di Caro, Daniel Ross (eds.),
Pseudo-coordination and Multiple Agreement Constructions,
Amsterdam / Philadelphia, John Benjamins, 2022**

1. Introduzione

Il volume scaturisce da due convegni dedicati alla pseudocoordinazione (Pse-Co) e alle costruzioni ad accordo multiplo (MAC) tenuti a Venezia nel 2017 e nel 2019. I 14 saggi che raccoglie toccano oltre 80 lingue, anche tipologicamente e geograficamente distanti (si veda l'indice delle pp. 337-338), in cui i fenomeni studiati sono documentati (in particolare, l'italiano e alcune varietà di Sicilia, Calabria e Puglia – capp. 2-5 –, il rumeno – cap. 6 –, il portoghese brasiliano e il polacco – cap. 7 –, il ceco – cap. 8 –, lo svedese – cap. 9 –, arabo, aramaico ed ebraico – cap. 10 –, il giapponese – cap. 11 – e il turco – cap. 12). I contributi sono dedicati a diversi aspetti delle due costruzioni (perlopiù negli ambiti della morfosintassi e della semantica), offrendo proposte differenziate e talvolta in opposizione le une alle altre. Il volume è aperto da un primo capitolo introduttivo, a firma della curatrice e dei curatori, “Pseudo-Coordination and Multiple Agreement Constructions: An overview”, cui seguono tre sezioni: “Romance languages” (capp. 2-6), “Other languages” (capp. 7-10) e “Comparative and theoretical” (capp. 11-14). Il volume (342 pp.) si chiude con un indice delle lingue e un indice degli argomenti.

2. Descrizione dei fenomeni e terminologia

I fenomeni cui il volume è dedicato sono stati notati e descritti in passato (già dal XVI sec., ma più compiutamente a partire da metà Ottocento) usando denominazioni diverse fra loro. Tale frammentazione terminologica, come mettono in evidenza Giusti, Di Caro e Ross (cap. 1, pp. 6-7), riflette la generale disconnessione tra tali descrizioni, che ha di fatto nascosto le non poche somiglianze tra le costruzioni presenti in lingue diverse. Di fatto, ad es., la pseudocoordinazione è stata perlopiù descritta in passato, dai singoli autori, come una caratteristica peculiare (o addirittura una «bizzarria») del codice indagato da ciascuno di loro. Uno degli obiettivi del volume è appunto pro-

muovere la consapevolezza che la pseudocoordinazione è invece un fenomeno cross-linguistico (*ibid.*).

2.1 Pseudocoordinazione

La pseudocoordinazione (nel volume *Pseudo-Coordination o pseudocoordinazione*, a seconda delle preferenze degli singoli autori; Cruschina, cap. 5, usa invece *Doubly Inflected Construction: DIC*), per la quale Giusti e Cardinaletti (cap. 2) adoperano l'abbreviazione PseCo (Di Caro 2017, 2019), è una costruzione che esibisce un unico soggetto e due verbi (V1 e V2) che concordano per tempo, aspetto, modo, persona e numero, separati solo da un connettore (spesso omofono di una congiunzione coordinante, talvolta assente/non pronunciato). La denominazione *pseudocoordinazione* è motivata dal fatto che i due verbi concordano esattamente come farebbero due verbi coordinati e dal fatto che il connettore è (omografo di) una congiunzione coordinante. Non si tratta però di una vera coordinazione perché, a differenza delle vere strutture coordinate, la pseudocoordinazione non consente l'inversione di ordine tra i due verbi (1), consente l'estrazione di un elemento interrogativo che sia argomento di V2 (2)-(3), e consente la procliticizzazione su V1 di un argomento di V2 (4)-(5):

- (1) a. *I'll go and get some milk* (pseudocoordinazione)
b. **I'll get some milk and go.* (cap. 1, es. (10)).
- (2) a. *Mary will come and visit them tomorrow* (pseudocoordinazione)
b. *Who will Mary come and visit __?* (cap. 1, es. (3))
- (3) a. *Mary went to her home town and visited her parents* (struttura coordinata)
b. **Who did Mary go to her home town and visit __?* (cap. 1, es. (2))
- (4) *Nun l'arrivà a ffici* 'non arrivò a farlo', lett. 'non lo arrivò a fece' (pseudocoordinazione, Deliano, cap. 4 es. (4)).
- (5) a. *Sono arrivato e ho fatto la prova* (struttura coordinata)
b. *Sono arrivato e l'ho fatta*
c. **La sono arrivato e ho fatto*

Il connettore nella PseCo di molti dialetti della Sicilia è *a* (e non, o non solo, *e*) e produce raddoppiamento fonosintattico (v. (4) *supra*). Tradizionalmente, tale *a* è considerato il continuatore del lat. AC (v. Di Caro, cap. 4, p. 100, nota 2), ma alcuni autori (ad es. Manzini e Lorusso, cap. 3) propongono che coincida

in sincronia con la preposizione *a* (< AD), usata come complementatore finito. Ciò potrebbe costituire un controesempio, secondo Giusti e Cardinaletti (cap. 2, p. 37), all'ipotesi che tali costruzioni siano pseudocoordinazioni. In realtà, però, come osserva anche Cruschina (cap. 5, pp. 130-131, nota 2), l'etimologia del connettore *a* è irrilevante ai fini dell'analisi sincronica, non solo perché tale elemento è semanticamente vacuo, ma anche, aggiungerei, perché la pseudo-coordinazione è definita, dalle stesse Giusti e Cardinaletti, appunto, come una 'falsa' coordinazione in sincronia (per via dell'accordo tra V1 e V2), e non necessariamente come una struttura che è o era davvero, in origine, coordinativa.

Quando il connettore è assente (6), la PseCo appare molto simile alle costruzioni *a verbo seriale* (*Serial Verb Construction: SVC*). Nella letteratura (v. cap. 1, p. 3), si è di fatto ipotizzato che le PseCo siano casi particolari di SVC, persino in presenza di un connettore (cap. 14, § 2).

- (6) *Vɔ 'mandʒə* 'vado a mangiare' (lett. 'vado mangio') [Martina Franca, cap. 2, p. 38, da Manzini e Savoia 2005: 690].

L'accostamento della PseCo alla SVC è però dovuto, come nota Ross (cap. 14, p. 318), al desiderio di connettere i singoli dati regionali ad un contesto tipologico più ampio, a costo di forzare la definizione di SVC.

Altre caratteristiche delle PseCo sono l'impossibilità di negare uno dei due verbi senza negare l'altro, il possibile sbiadimento semantico (e/o la riduzione da verbo a elemento (semi)funzionale) del V1 (che assume nella costruzione, ad es., valore aspettuale, che perde ogni significato, o che non ha una propria struttura argomentale autonoma da V2) (cap. 1, p. 11), la restrizione relativa ai lessemi verbali che possono realizzare il V1 (perlopiù verbi di movimento e di postura), e dei tempi e modi in cui la costruzione può realizzarsi, dando luogo a paradigmi difettivi (cap. 4).

2.2 Costruzione ad accordo multiplo (*Multiple Agreement Construction: MAC*)

Il termine *costruzione ad accordo multiplo* (*Multiple Agreement Construction: MAC*) è utilizzato nel volume con due significati, uno più generale e uno particolare. Nel significato generale, la MAC è una configurazione in cui i tratti di accordo verbale si propagano su più forme verbali anche se sono interpretabili solo su una di esse (in modo paragonabile a quanto avviene ai tratti di genere e numero

all'interno delle espressioni nominali romanze, cfr. cap. 1, p. 4; v. anche cap. 5, p. 132 nota 3). In tale accezione, MAC è un iperonimo di PseCo e SVC, che sono viste, cioè, come casi particolari di MAC (cap. 1, p. 3). Nell'accezione particolare, invece, con MAC si intendono quelle che Giusti e Cardinaletti (cap. 2) chiamano *muMAC*, vale a dire le costruzioni dell'Italia meridionale e dell'area messinese caratterizzate da un V1 e un V2 che concordano per persona e numero (ma con tempi indipendenti: V1 è libero, V2 è sempre e solo al presente indicativo), collegati da un continuatore di MODO ((*m*)*u*/*(m)**i*/*ma*) o di QUOD (*ku*):

- (7) *Vorria mu sacciu* 'vorrei sapere' (lett. 'vorrei *mu* so') [calabrese meridionale, cap. 1, p. 17 es. (28d), da Rohlf's 1969: 103].

Mentre nelle varietà che esibiscono la PseCo è possibile sostituire la sequenza connettore+V2 con *a* + infinito, la stessa sostituzione è invece generalmente impossibile nelle varietà caratterizzate dalla *muMAC*. Tale impossibilità è ricondotta da Rohlf's (1969, § 717) alla «impopolarità dell'infinito» nelle zone dell'Italia meridionale in cui il greco fu parlato fino al medioevo: la *muMAC* assume così il ruolo di strategia sintattica alternativa alle subordinate infinitive.

Gli altri aspetti per cui la *muMAC* si differenzia dalla PseCo sono la mancata concordanza per tempo e modo tra V1 e V2, cui si è già accennato, l'impossibilità della procliticizzazione su V1 di pronomi argomento di V2 (obbligatoria invece nella PseCo, cfr. (8) vs. (9)) e, cosa fondamentale, la possibilità di avere un soggetto distinto per V1 e V2 (10):

- (8) a. *Vinni mi ti vidu* 'sono venuto a vederti', lett. 'venni *mi* ti vedo' (Furci Siculo)
b. **Ti vinni mi vidu* lett. 'ti venni *mi* vedo'
- (9) a. **Vinni a ti vitti* lett. 'venni *a* ti vidi'
b. *Ti vinni a bbitti* 'ti sono venuto a vedere', lett. 'ti venni *a* vidi' (Deliano, cap. 1, p. 18)
- (10) *Vogghiu mi veni* 'voglio che venga.3SG', lett. 'voglio *mi* viene' (messinese, cap. 1, p. 4).

2.3 Pseudocoordinazione e pseudosubordinazione

Una classificazione in parte diversa è quella proposta nel lavoro di Edzard (cap. 10), basata su Yuasa e Sadock (2002) e ben illustrata con esempi tratti

da lingue semitiche. Si tratta di uno schema booleano che descrive le quattro possibili combinazioni dei valori ‘coordinazione’ / ‘subordinazione’ realizzati nella sintassi e/o nella semantica (cap. 10, p. 231):

	Sintassi	Semantica
Coordinazione “semplice”	coordinazione	coordinazione
Pseudocoordinazione	coordinazione	subordinazione
Subordinazione “semplice”	subordinazione	subordinazione
Pseudosubordinazione	subordinazione	coordinazione

Quelle che l’autore definisce coordinazione e subordinazione “semplici” sono caratterizzate quindi da una simmetria tra aspetto sintattico e aspetto semantico, quelle definite tramite il confisso *pseudo* sono invece costruzioni con una semantica che diverge da quanto ci si aspetterebbe in base alla struttura sintattica e alla sua interpretazione canonica. I casi di coordinazione sintattica includono sia la sovrabbondanza di congiunzioni, con congiunzione (correlativa) che precede anche il primo congiunto, sia, all’opposto, l’asindeto (è il caso delle SVC). Lo sbiadimento semantico (grammaticalizzazione/funzionalizzazione) del V1 nella pseudocoordinazione è una caratteristica ben documentata e che coinvolge più tipi di verbi negli esempi proposti da Edzard. Ad esempio, i V1 ‘aggiungere’ nell’ebraico biblico e ‘ritornare’ nell’accadico esprimono la reiterazione dell’atto denotato dal V2 (e il V1 può infatti essere glossato, come Edzard nota, con un avverbale come ‘di nuovo’):

(11) *way-yōsep ’abrāhām way-yiqqah ’iššā(h)* ‘E Abraham prese moglie di nuovo’
lett. ‘E aggiunse Abraham e-prese moglie’ (cap. 10, p. 233, da Gen. 25:1).

(12) *atūr-ma wardam ana bēli-ya aṭrud* ‘Mandai di nuovo lo schiavo al mio signore’
lett. ‘Ritornai-e lo schiavo al signore mandai’ (cap. 10, p. 233, da Huehnergard 2005: 125).

Similmente, il V1 ‘precedere’ nell’aramaico biblico assume un significato che può essere reso con l’avverbale ‘in anticipo’:

(13) *’aqdāmku wa-nāgārku-kamu* ‘Vi dissi in anticipo’ lett. ‘Precedetti e vi dissi’
(cap. 10, p. 233, 1 Thess 3:4).

Un grado maggiore di grammaticalizzazione è ben esemplificato dal verbo di postura ‘sedersi’, che esprime, in arabo levantino, come V1 della pseudocoordinazione, l’aspetto progressivo:

- (14) *qa’dat wa-katbat* ‘Stava scrivendo’ lett. ‘Si sedette e-scrisse’ (cap. 10, p. 232, da Ross 2016: 211).

3. Analisi formali

3.1 Pseudocoordinazione

Giusti e Cardinaletti (cap. 2: “Theory-driven approaches and empirical advances. A protocol for Pseudo-Coordinations and Multiple Agreement Constructions in Italo-Romance”) distinguono tra due sottotipi di pseudocoordinazione, la *ePseCo* e la *aPseCo*. Il primo è quello in cui il connettore tra V1 e V2 è omografo della congiunzione coordinante; il secondo è quello tipico dei dialetti della Sicilia e della Puglia, in cui il connettore è *a*.

3.1.1 *ePseCo*

Per la *ePseCo*, basandosi su Soto Gómez (2020), Giusti e Cardinaletti (cap. 2) propongono che V1 sia una testa funzionale della periferia sinistra (Focus) nella proiezione estesa di V2, per soddisfare le proprietà di selezione del connettore ‘e’; il connettore a sua volta riempie la posizione di complementatore, mentre il resto della frase è canonica, con V2 verbo finito della frase in posizione di complemento di C (11).

- (15) *ePseCo*: [_{FoCP} V1 [_{CP} ‘e’ [_{TP} V2 [_{VP} V₂ ... (cap. 2, p. 40 (8a))

Quindi, in una frase come *prendo e parto* (cfr. cap. 2 p. 37, es. (1)), *prendo* è un elemento funzionale in Focus ed *e* è un complementatore (non una congiunzione coordinante, quindi) che introduce la frase *parto*. Tra altre cose, quest’analisi spiega perché i clitici dipendenti da V2 non possono risalire su V1: i clitici risalgono verso la testa T più vicina e, secondo l’ipotesi in (15), nella *ePseCo* è presente un’unica testa T, su cui si risalda il V2, mentre V1 è generato come testa funzionale al di sopra dello strato TP.

Basandosi su de Vos (2005), Adina Camelia Bleotu (cap. 6: “The properties of the ‘(a) lua și X’ (‘take and X’) construction in Romanian. Evidence in favor

of a more fine-grained distinction among pseudocoordinative structures”) propone, per la costruzione ‘take and X’ che osserva in rumeno, accostabile alla *ePseCo*, un’analisi diversa, secondo cui V1 e V2 sono generati come elementi di una testa V complessa: il connettore lessicalizza una testa ‘take &’ (responsabile della specifica semantica del costrutto) che si salda al V2; il complesso ‘take &’ + V2 si salda poi al V1, per formare così V°, che si linearizza infine come V1+‘take&’+V2 (> V1+și ‘e’+V2) (cfr. *ivi*, pp. 158-159):

(16) ... [_v. V1 [‘e’ V2]]

Considerando l’assenza di gerarchia nella saldatura tra le teste ‘e’ e V2 (si comandano a vicenda), l’ordine nella linearizzazione delle due è stabilito solo da un principio di precedenza (da sinistra a destra), come in de Vos (2005), che però in termini più attuali risulta *ad hoc*. L’ipotesi in tal modo non sembra quindi in grado di impedire gli ordini agrammaticali, ad es. l’ordine *V1+V2+‘e’. Resta poco chiaro anche quale sia il filtro che impedisce la generazione indiscriminata di teste complesse come quella in questione (ad es. con V1 diversi da ‘prendere’, o prive del connettore, o del V2), o la saldatura di ulteriori teste alle tre contemplate in (16). Si potrebbe assumere che il V1 “funzionale” ‘prendere’ selezioni (e legittimi) la testa “take&” (in questo senso converrebbe che fosse “take&” a trasmettere la propria etichetta al complesso “take&”+V2, e non V2, cfr. p. 159, (18)) e che quest’ultima a sua volta richieda di saldarsi a un qualunque verbo V2. Bisognerebbe però assicurarsi a questo punto che le “vere” coordinazioni tra teste non siano generate mediante lo stesso tipo di derivazione (è proprio la distinzione dai costrutti coordinativi, infatti, a spiegare le altre proprietà del costrutto in esame, come la possibilità di estrazione da complementi di V2 ecc.). Quanto alla potenza descrittiva, l’idea della testa complessa spiega, secondo l’autrice, tra altre cose, l’impossibilità dell’inserimento di aggiunti tra il connettore e V2 (p. 153-154). Ciò è possibile a condizione, però, io credo, che si supponga che l’aggiunto non possa saldarsi a teste (semplici o complesse, come “take&+V2”), ma solo a sintagmi (che però sono anch’esse teste “complesse” nel modello *bare phrase structure* che sta alla base di de Vos 2005 e quindi dell’analisi di Bleotu). Un altro aspetto che rimane poco chiaro riguarda la proiezione degli argomenti di V1 e V2 dall’interno del complesso verbale.

Ancora un'analisi diversa è fornita da Mendes e Ruda (cap. 7: "Pseudo-coordination and ellipsis. Expressive insights from Brazilian Portuguese and Polish") per analoghe costruzioni (del tipo 'prendere e V') del portoghese brasiliano e del polacco. Secondo la loro analisi, V1 si salda al connettore 'e', formando un sintagma &P (di cui il connettore è la testa, ma sta a destra rispetto al complemento che contiene V1), che poi si salda come aggiunto a sinistra di un vP proiettato da V2 (ivi, pp. 180-181; 184):

- (17) ... [_{vP} [_{&P} [... V1] 'e'] [_{vP} ... V2 ...

L'analisi è motivata dal comportamento del V1 nelle risposte a eco alle interrogative polari e dalle similarità semantiche e sintattiche tra il complesso V1+connettore e gli elementi appositivi espressivi. Nello specifico, in portoghese brasiliano (ivi, p. 174), la risposta a un'interrogativa polare è ottenuta ripetendo il verbo lessicale principale della domanda (con ellissi di tutto il resto: A: *O João trouxe açúcar?* 'João ha portato lo zucchero?' B: *Trouxe 'si'*, lett. 'ha portato', ivi, p. 172). Se l'interrogativa è basata su una struttura ePseCo (ad es. *pegou e comprou* lett. 'prese e comprò', cfr. ivi, p. 174), risulta possibile rispondere ripetendo il verbo lessicale V2 (*comprou*), ma non il V1 (*pegou*):

- (18) A: *O João pegou e comprou café?* 'Giovanni ha comprato il caffè?', lett. 'Il João prese e comprò caffè?'.
B: *Comprou./ *Pegou 'si'*, lett. 'comprò/*prese'.

Il motivo di questa asimmetria, secondo Mendes e Ruda, è che il V1 (assieme al connettore) è collocato strutturalmente in una posizione che lo esclude, lo rende "invisibile", nel processo di formazione della risposta a eco. Lo stesso accade anche agli elementi espressivi, e in particolare alle apposizioni espressive:

- (19) A: *O João, idiota, finalmente chegou?* 'Quell'idiota di João finalmente è arrivato'
B: *Chegou, mas ele não é idiota* 'È arrivato, ma non è un idiota' (cap. 7, p. 179).

La possibilità in (19) di negare il solo contenuto dell'apposizione *idiota* indica che tale apposizione non è inclusa nella parte ellittica della risposta a eco. Mendes e Ruda propongono dunque che il complesso V1+connettore della pseudocoordinazione appartenga all'«expressive realm of language» e occupi la stessa po-

sizione appunto delle apposizioni espressive. Basandosi su Biberauer e Vikner (2017), collocano tali elementi espressivi sul margine di una fase, come aggiunti. In questo caso, scelgono l'aggiunzione a vP, vale a dire il sintagma verbale proiettato dal verbo lessicale V2 (17). La possibilità di avere la negazione e alcuni avverbi tra il connettore e il V2 (*pegou e não/finalmente comprou* 'prese e non/finalmente comprò' vs. **não pegou e comprou*, cfr. p. 183), corrobora, secondo gli autori, l'ipotesi che il connettore formi un costituente con V1 e non con V2.

L'analisi è estremamente interessante, perché tenta di spiegare in modo unificato due fenomeni apparentemente non correlati, la pseudocoordinazione e gli espressivi. Tuttavia, l'aggiunzione a sinistra del complesso V1+connettore e, al suo interno, la scelta di avere V1 a sinistra del connettore, appaiono soluzioni *ad hoc*. Bisognerebbe infatti chiarire cosa impedisca al connettore di essere linearizzato a sinistra di V1 (**... e pegou comprou...*), considerando tra l'altro che, nella rappresentazione data dagli autori, il connettore è la testa di &P (e il brasiliano è una lingua con testa a sinistra). Allo stesso modo, dato che i due vP si comandano a vicenda, manca un meccanismo che impedisca che l'aggiunto V1+connettore venga linearizzato (anche, opzionalmente) a destra del vP proiettato da V2, facendo così finire la stringa frasale con V1+connettore (o connettore+V1) (**comprou o café pegou e/*comprou o café e pegou*).

Anche il costrutto indagato da Shimada e Nagano in due varietà giapponesi (cap. 11: "Ambiguities in Japanese pseudo-coordination and its dialectal variation") può essere accostato alla ePseCo. Nella loro analisi, l'ordine superficiale V1-*te* V2 è ottenuto tramite la canonica encliticizzazione della congiunzione *-te* su V1, quindi sul margine destro del sintagma (il giapponese è una lingua con testa a destra); il costituente così ottenuto viene poi selezionato da una preposizione non pronunciata; il sintagma che ne risulta è infine selezionato dal V2 *iru/aru* 'esistere', che si linearizza così ancora più a destra:

(20) [... [_{pp} [... V1-*te*] Ø] V2] (cfr. *ivi*, p. 259)

A dispetto del *-te*, dunque, non c'è vera coordinazione, né sintatticamente né semanticamente: il V2 aggiunge al predicato espresso da V1 una particolare coloritura (esistenziale, progressiva ecc., ad es. *kuruma-ga tome-te aru* lett. 'automobile-NOM parcheggiare-*te* esistere', propr. 'c'è un'automobile parcheggiata', *ibid.*;

John-ga ronbun-o kai-te iru lett. ‘John-NOM saggio-ACC scrivere-*te* esistere’, propr. ‘John sta scrivendo/ha completato un articolo’, cap. 11, p. 247). L’autore e l’autrice si dedicano principalmente a spiegare alcune ambiguità del costrutto (che riconducono principalmente a differenze nella derivazione), senza approfondire la riflessione sulla saldatura del V2 e sul suo contributo semantico, la cui natura però è sottesa alla struttura formale proposta, basata su Mihara (1997) e Matsuoka (2019). Cosa impedisce a *-te* di encliticizzarsi su qualunque verbo in assenza di un secondo congiunto con verbi sovraordinati diversi da *aru/iru*? *Aru/iru*, nella specifica accezione/funzione qui in esame, seleziona esclusivamente il PP che ha per testa questa particolare preposizione \emptyset , che a sua volta seleziona esclusivamente un costituente che si chiude con il *te* “non coordinante”. Specularmente, ciascuno degli elementi selezionati non può essere selezionato/legittimato da teste diverse da quelle già menzionate. Insomma, una serie di selezioni “idiosincratiche” a cascata tipica delle proiezioni funzionali. Shimada e Nagano, riprendendo Matsuoka, etichettano il complemento di *aru/iru* come PP (con testa una preposizione \emptyset), anche se al suo interno sono presenti un nominale (realizzato in alcuni casi da PRO) e il predicato V1. Tale notazione PP è così di fatto equivalente a CP (con complementatore \emptyset). Resta comunque poco chiaro, perlomeno dalla sola lettura di questo saggio di Shimada e Nagano, l’utilità della presenza della testa \emptyset e della sua proiezione. Senza tale testa \emptyset , si potrebbe ipotizzare che *aru/iru* selezioni direttamente un costituente la cui testa sia *-te*, ottenendo così una struttura congruente con quella ipotizzata da Soto Gómez per lo spagnolo e adottata da Giusti e Cardinaletti per le ePseCo (15). Poiché il giapponese ha la testa a destra, il V1 corrisponderebbe al V2 di (15), il *-te* alla testa C e il V2 al V1, che è infatti il verbo meno incassato e di natura (semi)funzionale.

Mitrović (cap. 13: “Syntax, semantics, and pragmatics of Pseudo-Coordination”) si occupa dei costrutti del tipo *go-(and-)get* in inglese, tramite un’analisi che parte dalla semantica formale. Contrariamente alle altre proposte fin qui elencate, Mitrović preserva in qualche modo, (anche) in sincronia, la natura coordinativa della pseudocoordinazione da un punto di vista formale. Basandosi su Mitrović (2014 e 2021), propone infatti che la PseCo si opponga alla coordinazione standard perché nella prima, a differenza che nella seconda, i due elementi congiunti sono “asimmetrici”, ma condividono comunque

una serie di tratti essenziali che consentono alla derivazione di non interrompersi. La PseCo è considerata così una «improper junction» derivata attraverso un processo che l'autore definisce «Dynamic Conjunction». Tra le altre cose, l'ipotesi rende conto anche della derivazione della funzione di sorpresa della PseCo, studiata anche da Cruschina per le aPseCo (cap. 5, v. *infra*).

Škodová (cap. 8: “Pseudo-coordination of the verb *jít* ('go') in contemporary Czech”), studiando la ePseCo che coinvolge il V1 *jít* ‘andare a piedi’ in ceco, sembra voler preservare, come Mitrović, perlomeno dal punto di vista interpretativo, la natura di coordinatore del connettore *a* ‘e’ (omofono della congiunzione coordinante in ceco), che, secondo la sua analisi, “congiunge” due micro-stadi («sub-stages») di un unico evento complesso espresso dalla combinazione di V1 e V2. L'autrice afferma che il primo verbo denota cioè la fase preparatoria dell'attività, il cui nucleo è invece espresso dal secondo verbo. In realtà, dalla lettura del saggio, mi sembra che tale analisi possa applicarsi forse solo ai casi in cui V1 ‘andare’ mantiene il significato lessicale di movimento, nel senso che l'azione di ‘andare a fare X’, ad es., è vista come un'unica azione che inizia con uno spostamento (la natura monoeventiva del costruito viene testata in modo convincente attraverso una serie di test). Ci sono altri casi però in cui l'autrice mostra che lo stesso V1 contribuisce in modo diverso alla semantica dell'enunciato, esprimendo ad es. l'avvio improvviso dell'evento descritto da V2 (ivi, p. 196). In questi casi, V1 ha quindi valori aspettuali o modali che non possono essere ricondotti a una fase, a uno stadio dell'evento, ma, appunto, semmai, a una sua caratteristica.

All'ambito della ePseCo è possibile ricondurre anche le costruzioni studiate da Blensienius e Andersson Lilja nello svedese (cap. 9: “In search of subjective meaning in Swedish pseudocoordination”). Nel loro saggio l'aspetto principalmente indagato è quello semantico, in una prospettiva anche diacronica. Più specificamente, i due autori tentano di ricostruire i percorsi del mutamento semantico dei V1 ‘andare’ e ‘sedersi’, che perdono il significato lessicale originario e contribuiscono, nella pseudocoordinazione, ad esprimere vari contenuti soggettivi, secondo un quadro che può essere accostato a quello descritto anche da Cruschina per le aPseCo (v. *infra*).

3.1.2 *aPseCo*

Per quanto riguarda la *aPseCo*, Giusti e Cardinaletti (cap. 2, p. 40) (rifacendosi a Cardinaletti e Giusti 2020) propongono che V1 sia la lessicalizzazione di una testa funzionale *t*, una reiterazione della categoria T (tempo). Il connettore *a*, che in Cardinaletti e Giusti (2020: 127) era collocato in un costituente privo di etichetta posto tra *tP* e TP, figura adesso all'interno di un'ulteriore proiezione *tP*, intermedia tra il *tP* più alto del V1 e il TP che ospita V2:

$$(21) \text{ } aPseCo: [_{tP} V1 [_{tP} (a) [_{TP} V2 [_{vP} \forall Z \dots \text{ (cfr. cap. 2, p. 40 (8b))}$$

Le autrici non approfondiscono l'analisi di tale connettore. La notazione riportata in (21) lascerebbe supporre che suggeriscano si tratti di una testa, che ha una versione con significante \emptyset . Ma non è possibile escludere che l'assenza del connettore dalla stringa fonetica non implichi invece anche l'assenza dello strato che la contiene, tenendo conto del fatto che nel citato articolo del 2020 le autrici distinguono tra i casi in cui il connettore non è distinguibile in superficie ma produce raddoppiamento sintattico della consonante iniziale del V2 (es. *lu va ffazzu* lett. 'lo vado faccio') e i casi in cui manca anche il raddoppiamento sintattico (es. *vaju pigliu u pani* lett. 'vado prendo il pane'), e il connettore viene allora definito «truly missing» (Cardinaletti e Giusti 2020: 127).

L'analisi in (21) rende conto di tutte le proprietà della costruzione. Innanzitutto, la natura funzionale di V1 (come anche nella *ePseCo*) predice che tale verbo appartenga a una classe chiusa e che non proietti una struttura argomentale (come gli ausiliari, cfr. cap. 2, p. 41). La natura "posticcia" («fake») dei tratti di tempo e aspetto di *t* è inoltre coerente con la difettività della flessione di V1 o la sua realizzazione ridotta in alcune varietà. Lo strato *t/TP* predice infine la risalita a sinistra di V1 della negazione e dei clitici argomento di V2 e spiega l'adiacenza tra V1, connettore e V2.

La parte meno sviluppata è quella che riguarda il carattere e la posizione del connettore. Se, infatti, considerare V1 come la lessicalizzazione di una testa flessiva (*t*) è coerente con l'analisi tradizionale di altri elementi funzionali flessivi, come gli ausiliari, lo stesso non si può dire per il connettore *a*, che pure, in questa versione della teoria, sembra lessicalizzare un'istanza della stessa testa *t*. La scelta sembra determinata dalla necessità di non avere categorie distinte che separino TP da *tP*. Tuttavia, è convincente l'idea espressa alle pp. 39-40 che

l'assenza della *aPseCo* nell'italiano standard (e in altre varietà) si spieghi facilmente postulando l'assenza dal lessico di questi codici di un item a_1 (\emptyset) totalmente distinto dal più diffuso complementatore a_2 delle infinitive (aggiungo io gli indici per comodità), e appartenente ad una categoria chiaramente diversa (*t* vs. *C*). a_1 è dunque caratterizzato da un corredo in tratti che lo distingue da a_2 , ma che lo accomuna ai verbi *V1*, cioè quelli ammessi al first merge in t^* . Sempre da tale corredo si può far dipendere la selezione di *TP* e non di *tP* (per filtrare strutture ricorsive e ordini agrammaticali come *... *a a* ..., **V1 a V1*... ecc.). Qualunque sia l'origine storica di a_1 (da *AC*, *AD*, o altro), il costruito in questione sembra modellato da processi diacronici di rianalisi (ricategorizzazione) che investono la flessione e la struttura della frase e che potrebbero motivare gli aspetti idiosincratici che abbiamo evidenziato, in particolare la reiterazione di *t* e il corredo in tratti del connettore, che esibisce probabilmente caratteristiche relative alla nuova categoria assieme a residui della vecchia (o delle vecchie).

All'analisi formale di Giusti e Cardinaletti si oppone diametralmente quella di Manzini e Lorusso (cap. 4: "A bisentential syntax for *a*/*bare* finite complements in South Italian varieties. Motion verbs and the progressive"), che propongono per la *aPseCo* una struttura bifrasale (non dissimile da quella delle frasi con subordinata infinitiva e con *muMAC*) e l'identificazione del connettore della *aPseCo* con il complementatore *a* che introduce subordinate infinitive. La loro analisi punta all'economia sia per quanto riguarda la struttura che per quanto riguarda il lessico: non vengono ipotizzate strutture innovative (v. il *tP* di Giusti e Cardinaletti), il connettore è una preposizione con funzione di complementatore, non distinta, nel lessico, da quella che introduce le infinitive. Le peculiari proprietà che oppongono la *aPseCo* alle *muMAC* e alle frasi con subordinata infinitiva e di modo finito, nella loro proposta, sono fatte dipendere, quindi, necessariamente, da altri fattori e non da caratteristiche strettamente strutturali.

Innanzitutto, la presenza del complementatore *a* viene associata non alla flessione non finita della subordinata (come nella tradizione cartografia a partire da Rizzi 1997), ma al fatto che verbo sovraordinato e subordinato non hanno specificazioni di tempo distinte e indipendenti. Così, *a* può introdurre, ad es., le infinitive, perché queste ultime non hanno una specificazione temporale distinta da quella del verbo reggente. Similmente, la frase subordinata proiettata da

V2 nella α PseCo è introdotta da *a* perché o V2 è privo di una vera specificazione temporale (come gli infiniti) o concorda in tempo con V1 (cap. 3, p. 77), e dunque non ha, in nessun caso, una specificazione temporale distinta e indipendente.

Per quanto riguarda la risalita su V1 dei clitici dipendenti da V2, obbligatoria nella α PseCo ma impossibile nelle *mu*MAC e in altre strutture bifrasali, Manzini e Lorusso ipotizzano (basandosi su Kayne 2000) che il fenomeno dipenda non dalla presenza/assenza dello strato CP della subordinata, ma dalla natura “fasale”/non “fasale” della testa C ovvero dalla sua “difettività” (cap 3, p. 84). La risalita è cioè possibile solo se la testa C è difettiva/non fasale. Assumono, quindi, che nelle α PseCo manchi una testa C fasale che impedisca il movimento lungo del clitico. Un problema che Manzini e Lorusso riconoscono è che ciò non basterebbe a rendere conto dell’obbligatorietà della risalita nelle α PseCo, dato che la testa C difettiva consente ma non forza la risalita. Propongono ulteriormente, allora, che esistano vari gradi di difettività di C, da cui potrebbe dipendere appunto la differenza tra facoltatività e obbligatorietà della risalita (v. *ibid.*).

Non mi pare che sia spiegata l’impossibilità di avere soggetti distinti per V1 e V2, come invece ci si aspetterebbe in una struttura bifrasale. Allo stesso modo mancano elementi che possano spiegare il numero ristretto di verbi ammessi nella posizione di V1 e la difettività della loro flessione in alcune varietà.

La monofrasalità (intesa anche semanticamente, come ‘co-predicazione’ ovvero mutua dipendenza tra le azioni espresse dai due verbi, v. anche *supra* l’analisi di Škodová) è fondamentale per la formulazione, da parte di Ross (cap. 14: “Pseudocoordination and Serial Verb Constructions as Multi-Verb Predicates”), di una categorizzazione diversamente articolata. Partendo dalle sue ricerche tipologiche sulla pseudocoordinazione e sulle SVC (Ross 2016, 2021), lo studioso propone l’introduzione di una nuova categoria, che chiama *Multi-Verb Predicates* (MVP) che comprende le SVC e le PseCo. La categoria del MVP permette così di evidenziare i punti di contatto tra PseCo e SVC senza forzare la definizione tradizionale di quest’ultima per includervi esempi della prima (v. anche *supra*). Nello specifico, i MVP sono costruzioni multi-verbo di natura monofrasale (i due verbi formano un predicato “complesso”). Tale monofrasalità spiega la condivisione di tratti di accordo tra i due verbi. All’interno della categoria generale dei MVP, la PseCo presenta un

connettore, mentre le SVC ne sono prive. L'aspetto decisivo è però la semantica e non la struttura. Sono infatti escluse dai MVP, ad es., le costruzioni che pure presentano un connettore omofono di 'e', ma che non sono monofrasali perché V1 e V2 non sono «co-predicati», ovvero non c'è 'mutua dipendenza semantica' tra le azioni espresse da V1 e V2. Così, ad es. l'autore oppone *I will try and finish the report on time, but I might not succeed* (MVP) a *I will go and get the book (#even if it is sold out)* (cap. 4, p. 325), o il brindisino *vogghiu a bbesciu* 'voglio uscire' (lett. 'voglio a esco') (ivi, p. 329) (entrambe strutture con 'complementazione' e quindi non appartenenti ai MVP).

Su un aspetto semantico e pragmatico della *aPseCo* si concentra infine Cruschina (cap. 5: "Gone unexpectedly. Pseudo-coordination and the expression of surprise"). Lo studioso analizza il V1 'andare' nella *aPseCo* in siciliano, evidenziando che il contributo interpretativo di tale V1 consiste, in una delle sue accezioni, nella valutazione dell'evento espresso dal V2 come inatteso da parte del mittente (ad es. *Vaiu a ssientu ca iddu ci fici stu tuortu a sso mughieri* 'ho sentito (lett. 'vado a sento') che lui ha fatto questo torto a sua moglie', p. 133, da Sornicola 1976). L'autore riconduce diacronicamente la formazione di tale implicatura convenzionale alla grammaticalizzazione cui i verbi di movimento sono andati incontro in diverse lingue, assumendo ad es. la funzione di marche di aspetto o tempo. Nello specifico, il passaggio dalla semantica lessicale del movimento all'espressione della sorpresa sarebbe stato generato metaforicamente: dall'andare via allontanandosi da un luogo fisico all'"allontanamento" figurato dell'evento riferito rispetto alle aspettative del mittente.

3.2 *muMAC*

Per la *muMAC*, Giusti e Cardinaletti (cap. 2) propongono una struttura bifrasale, che si oppone a quelle di *ePseCo* ed *aPseCo*, che le studiose analizzano invece come monofrasali: nella *muMAC*, V1 è un verbo sovraordinato che sottocategorizza una subordinata "ridotta" (cioè priva delle proiezioni funzionali più alte, come topic e focus), il cui complementatore (*mu/ku* ecc.) occupa la posizione di testa di FinP, la stessa, cioè, dei complementatori non finiti ('a', 'di', 'per'), a dispetto della flessione finita della subordinata:

- (22) *muMAC*: [_{TP} V1 [_{VP} $\forall \pm$ [_{FinP} (*mu/ku*) [_{TP} V2 [_{VP} $\forall \pm$... (cfr. cap. 2, p. 40 (8c))

L'analisi in (22) rende conto dell'impossibilità della risalita su V1 dei clitici dipendenti da V2, perché i clitici si fermano sul T più basso, dove V2 viene risaldato. La natura "difettiva" della proiezione FinP motiva la morfologia ridotta del V2 (solo all'indicativo presente, *ivi*, p. 41). Poiché sono presenti due proiezioni frasali, è possibile avere due soggetti distinti, mentre lo stesso è impossibile nelle PseCo, che hanno invece, secondo l'analisi delle due studiose, strutture monofrasali.

4. Altri aspetti

Prendendo le mosse dai suoi studi precedenti sulla PseCo, Di Caro nel cap. 4 ("Preterite indicative Pseudo-Coordination and morphomic patterns. The case of the W-Pattern in the dialect of Delia") si concentra sulla morfologia flessionale della *aPseCo* in siciliano. Lo studio, alla luce dei giudizi di 140 parlanti della varietà di Delia, evidenzia non solo l'esistenza di un peculiare «Pattern W», per cui la *aPseCo* nel passato remoto è possibile in tutte le persone ad esclusione della seconda singolare e della seconda plurale, ma conferma anche che nelle altre persone la costruzione è consentita solo se si selezionano forme di V2 con radici perfettive (rizotoniche) (*ivi*, p. 109). Il contrasto può riguardare coppie di verbi sinonimi, come ad es. *ntìsi vs. sintì* 'sentì, udì' (23), così come anche, in modo cruciale, coppie di varianti appartenenti allo stesso lessema, come *mòrsi vs. murì* 'morì' (24):

(23) *Tutta a nna vota, lu ji a 'ntisi/* sintì* 'All'improvviso lo sentì' (lett. 'All'improvviso lo andò a sentì')

(24) a. *Ji a mmorsi propriu oi* '(Tra tutti i giorni in cui poteva morire) è morto proprio oggi' (lett. 'andò a morì proprio oggi').

b. **Ji a mmurì propriu oi.*

Il cap. 12, di Tat e Kornfilt ("Partial versus full agreement in Turkish possessive and clausal DP-Coordination"), infine, non tratta direttamente le costruzioni verbali della PseCo e della MAC, ma si dedica ad uno degli aspetti costitutivi di tali costruzioni, vale a dire l'espressione "inattesa" di tratti di accordo su una testa. Il dominio indagato non è quello verbale ma quello nominale, la cui struttura interna è però simile a quella del dominio verbale, sebbene semplificata, secondo molti studiosi. Nello specifico, Tat e Kornfilt si concentrano sull'accordo parziale

nei sintagmi nominali possessivi in turco con predicati nominalizzati. Ipotizzano che tale accordo parziale sia il risultato non solo di processi interni al componente sintattico, ma all'intervento di processi che hanno luogo in un separato componente post-sintattico e che alterano il risultato del componente sintattico.

5. Conclusioni

Il volume raccoglie contributi che esibiscono punti di vista diversi e spesso in opposizione gli uni agli altri, in un dibattito vivace e libero che è però la condizione essenziale per il progresso della conoscenza. Tale disomogeneità si riflette nell'assenza di una terminologia unificata, che è in parte compensata dai non pochi rimandi interni e dai percorsi di lettura suggeriti dal capitolo introduttivo e dall'indice analitico. Le difficoltà che si rilevano in talune analisi non inficiano la solidità degli impianti metodologici, il rigore con cui le ricerche sono condotte, e il valore di ciascuna delle intuizioni sottese a tali ricerche. Anzi, la lettura del volume è certamente preziosa per una vasta platea di linguisti e spinge a interrogarsi e sui costrutti specificamente studiati e su questioni più generali, come la struttura della frase, l'accordo, i fenomeni di interfaccia tra i componenti della grammatica, la natura delle categorizzazioni nel confronto tra lingue tipologicamente distanti e il mutamento linguistico, con particolare attenzione alla grammaticalizzazione.

Bibliografia citata

- Biberauer, T., Vikner, S., 2017, "Having the edge: A new perspective on pseudo-coordination in Danish and Afrikaans", in N. LaCara, K. Moulton, A.-M. Tessier (eds), *A schrift to fest Kyle Johnson*, Linguistics Open Access Publications, 1, 77-90.
- Cardinaletti, A., Giusti, G., 2020, "Multiple agreement in southern Italian dialects", in L. Franco, P. Lorusso (eds.), *Linguistic variation: structure and interpretation*, Berlin, de Gruyter Mouton, 125-148.
- Di Caro, V.N., 2017, "Multiple Agreement Constructions: A Macro-Comparative Analysis of Pseudo-Coordination with the Motion Verb Go in the Arabic and Sicilian Dialects", in *Bucharest Working Papers in Linguistics* 19(2).
- Di Caro, V.N., 2019, *Multiple Agreement Constructions in Southern Italo-Romance. The Syntax of Sicilian Pseudo-Coordination*. PhD dissertation, Ca' Foscari University of Venice.

- Huehnergard, J., 2005, *A Grammar of Akkadian*, 2nd edition, Winona Lake, Eisenbrauns.
- Kayne, R., 2000, *Parameters and Universals*, Oxford, OUP.
- Manzini, R., Savoia, L., 2005, *I dialetti italiani e romanci: morfosintassi*, vol. 1, Alessandria, Edizioni dell'Orso.
- Matsuoka, M., 2019, “‘teiru’ shinkoobun no toogokoozoo to suuryoofukushi no kai-shaku ni tsuite [syntactic structures of ‘teiru’ progressives and the interpretation of adverbial quantifiers]”, in K. Takezawa et al. (eds.), *Nihongo Toogoronkenkyuu no Hirogari [Toward an Integrated Approach to Japanese Syntax]*, Tokyo, Kurosio, 25-44.
- Mihara, K., 1997, “Dooshi no asupekutokoozoo [Aspect of verbs]”, in K. Mihara, R. Washio (eds.), *Voisu to Asupekuto [Voice and Aspect]*, Tokyo, Kenkyusha, 107-196.
- Mitrović, M., 2014, *Morphosyntactic atoms of propositional logic: a philological programme*, University of Cambridge dissertation.
- Mitrović, M., 2021, *Superparticles: A Microsemantic Theory, Typology, and History of Logical Atoms*, Dordrecht, Springer.
- Rizzi, L., 1997, “The Fine Structure of the Left Periphery”, in L. Haegeman (ed.), *Elements of Grammar*, Dordrecht, Kluwer, 281-337.
- Rohlf, G., 1969, *Grammatica storica della lingua italiana e dei suoi dialetti*. Vol. 3: *Sintassi e formazione delle parole*, Torino, Einaudi.
- Ross, D., 2016, “Between coordination and subordination: typological, structural and diachronic perspectives on pseudocoordination”, in F. Pratas, S. Pereira, C. Pinto (eds.), *Coordination and Subordination: Form and Meaning – Selected Papers from CSI Lisbon 2014*, 209-243, Newcastle upon Tyne, Cambridge Scholars Publishing.
- Ross, D., 2021, *Pseudocoordination, Serial Verb Constructions and Multi-Verb Predicates: The relationship between form and structure*, Ph.D. dissertation, University of Illinois UrbanaChampaign.
- Sornicola, R., 1976, “Vado a dire, vaiu a ddicu: problema sintattico o problema semantico?”, *Lingua Nostra*, 37(3-4), 65-74.
- Soto Gómez, J.F., 2020, *Pseudocoordination in Spanish. A two construction analysis*, MA thesis, Ca' Foscari University of Venice. <http://dspace.unive.it/handle/10579/18854>.
- de Vos, M.A., 2005, *The syntax of verbal pseudo-coordination in English and Afrikaans*, Utrecht, LOT.
- Yuasa, E., Sadock, J.M., 2002, “Pseudo-subordination: a mismatch between syntax and semantics”, *Journal of Linguistics*, 38, 87-111.

Salvatore Menza

**Elisabetta Bonvino, Diego Cortés Velasquez,
Anna De Meo, Elisa Fiorenza, *Agire in L2.*
Processi e strumenti nella linguistica educativa,
Milano, Hoepli, 2023**

Agire in L2. Processi e strumenti nella linguistica educativa è un'ouverture di avvio allo studio della linguistica educativa in cinque atti. Il volume, curato da Elisabetta Bonvino, Diego Cortés Velasquez, Anna De Meo ed Elisa Fiorenza, intercetta e approfondisce i processi e gli strumenti dell'agire linguistico dalla lingua materna alla lingua seconda. Il libro consta di cinque capitoli, ognuno dedicato all'approfondimento di una abilità linguistica: il primo introduttivo, di tipo concettuale e terminologico; il secondo analitico e riflessivo, incentrato sull'abilità dell'ascolto; il terzo comparatistico, sul parlato; il quarto descrittivo, sulla scrittura; il quinto situazionale, sulla lettura.

Il primo capitolo «Coordinate» dota il lettore di strumenti guida per orientarsi nel mondo della didattica delle lingue e adotta la prospettiva della linguistica educativa per descrivere i profili degli apprendenti, dei parlanti plurilingue e del repertorio linguistico, secondo la dimensione temporale (plurilinguismo precoce o tardivo), la modalità di apprendimento (spontaneo, guidato o misto) e il contesto di esposizione alla lingua (di lingua seconda o L2, di lingua straniera o LS). Tali coordinate definiscono i fenomeni del multilinguismo e plurilinguismo, del monolinguisimo istituzionale (completo o parziale) e del bilinguismo (precoce, simultaneo, compatto, bilanciato, verticale, ricettivo) e distinguono le abilità primarie dalle abilità integrate nei processi di produzione e ricezione, secondo il canale di trasmissione (fonico-uditivo, gestuale-visivo, grafico-visivo) e secondo le modalità di comunicazione tra i parlanti (sincrona o asincrona). Si inquadrano quindi le caratteristiche e gli stadi dell'interlingua, le tipologie di errore dell'apprendente, il transfer (positivo e negativo), la consapevolezza metalinguistica e si approfondiscono i fattori di promozione dell'apprendimento, le fasi di elaborazione le caratteristiche e le possibili modifiche dell'input, gli interventi di potenziamento acquisizionale e gli eventi di negoziazione del significato e della forma.

Il secondo capitolo «Ascoltare» considera i processi di sviluppo dell'abilità nelle tre fasi (l'udito, la decodifica, la comprensione), dell'ascolto linguistico in L2 e studia la genesi e propagazione del suono nelle sue componenti costitutive, con affondi nell'ambito della fisica, della musica e della neuro-linguistica. L'analisi coinvolge la dimensione percettiva, i processi *top-down* e *bottom-up*, opera una distinzione per tipologia d'ascolto (estensivo, attento, interattivo, intensivo, selettivo, in attesa) e presenta una serie di tecniche e attività per esercitare e verificare la padronanza dell'abilità.

Il terzo capitolo «Parlare» indaga il ciclo evolutivo dell'abilità comunicativa a partire dal processo di fonazione, indagato dal punto di vista fisiologico, prosodico, vocale e articolatorio. La produzione linguistica si configura in parallelo per analogie e differenze di condizione e modalità, tra lo scritto e il parlato, secondo la materia, il canale, il codice, la dipendenza dal contesto, la pianificazione, la durata e il grado di dinamicità.

Il volume ha il pregio di affiancare agli studi scientifici frequenti riproduzioni figurative delle aree cerebrali interessate dai processi linguistici, spettrogrammi rappresentativi dei fenomeni articolatori a livello soprasegmentale ed esempi di produzione interlinguistiche, analizzati nella progressività dell'enunciazione, in base alla modalità di produzione, alle caratteristiche dell'interlingua e alle strategie comunicative (di conseguimento o di elusione) impiegate.

Il quarto capitolo «Scrivere» propone un percorso sulla scrittura come prodotto e codice, investiga la sua distribuzione diatopica e diacronica, ne ricostruisce l'origine come sistema alfabetico, con cenni alla fonetica e alla linguistica storica, dai sistemi fonologici e non fonologici alla didattica della scrittura. Emergono le potenzialità visive e rappresentative della comunicazione scritta. Con dovizia di esempi e raffronti si esaminano indicatori (coerenza, coesione, continuità, progressione tematica), parametri (ricorsività, simultaneità), fasi e peculiarità del processo di scrittura, in rapporto alle altre abilità, tra L1 e L2. Di particolare rilievo la sezione conclusiva sulla didattica della scrittura in L2, per le attualizzazioni, per gli orientamenti di ricerca e per le considerazioni chiave intorno al feedback correttivo.

Il quinto capitolo «Leggere» indaga strategie di apprendimento e benefici socio-cognitivi di potenziamento dell'abilità, con scenari situazionali identi-

ficativi del processo di lettura e delle implicazioni sul piano della conoscenza, della percezione e della valutazione in rapporto alle altre abilità primarie. Si invita a riflettere sulla consecutività delle tre fasi di lettura, di percezione del testo scritto, di decodifica (fonologica e lessicale) e di comprensione, con interessanti osservazioni sulle potenzialità e i limiti del testo digitale.

Il volume si conclude con una nutrita bibliografia e sitografia, a cui si aggiungono le espansioni online, un prezioso repertorio di risorse educative accessibili dalla piattaforma dedicata.

Se ne consiglia la lettura a docenti e studenti universitari, a educatori e formatori, per il rigore dell'analisi scientifica, per l'intelligibilità e la chiarezza del linguaggio, per la dimensione dialogica ed esemplificativa della narrazione, arricchita da numerose attività comunicative calate nei contesti d'uso della lingua, corredate da direttrici operative e da schede dettagliate dei fenomeni linguistici per adattarsi a una pluralità di applicazioni e interventi didattici.

Valentina Paganini

**Andrea Moro, *La razza e la lingua. Sei lezioni sul razzismo*,
Milano, La nave di Teseo editore, 2019**

In questo volume, Moro affronta la tematica del razzismo linguistico contestualizzando e discutendo quelli che egli ne ritiene essere i due presupposti essenziali: da un lato, l'idea che esistano lingue migliori di altre; dall'altro, l'idea secondo cui a lingue diverse corrisponderebbero modi diversi di vedere la realtà, e quindi di pensare. L'opera si divide in sei capitoli, intitolati rispettivamente *Che razza di lingua* (cap. I), *Perché non accada più (in linguistica)* (cap. II), *I confini di Babele ovvero una soluzione inaspettata* (cap. III), *Parole, sensazioni e ragioni* (cap. IV), *Dove vanno le lingue* (cap. V) e *Osservazioni conclusive* (cap. VI). Completano il volume un'appendice, intitolata *Una lettera di Cartesio sul linguaggio*, la bibliografia e i ringraziamenti.

Moro sgombra il campo da eventuali fraintendimenti fin dalla prima pagina del primo capitolo: secondo l'autore, il pericolo di un ritorno del razzismo si concretizza non tanto nell'uso improprio della parola "razza", bensì nel permanere nel pensiero comune dell'assunto, solo apparentemente innocuo, secondo cui le lingue si differenzierebbero le une dalle altre per un insieme di supposte caratteristiche per lo più arbitrarie (come ad esempio la complessità, l'ampiezza lessicale, la trasparenza o la capacità di favorire il ragionamento) che le renderebbero più o meno funzionali a compiti «di tipo cognitivo, culturale, psicologico o comunicativo» (p. 19). Espungere dal dizionario il termine "razza", magari sostituendolo, come proposto da molti, con il meno connotato ideologicamente "etnia", non solo non risolverebbe la questione, sostiene Moro, anzi ne ritarderebbe oltremodo la soluzione: servirebbe soltanto a darci l'illusione di aver eliminato l'idea di una disparità dal nostro modo di pensare, distogliendoci tuttavia dal dovere di prendere un'iniziativa nel concreto. E non sarebbe utile nemmeno negare, continua l'autore, che determinate caratteristiche fisiche a torto o a ragione associate a determinate "razze" «siano funzionali a un certo compito» (pp. 17-18), come ad esempio l'altezza media in ambito sportivo. Ciò che invece è imprescindibile eliminare è l'idea che esistano, accanto a differenze di carattere somatico e quindi meramente superficiali, delle differenze di carattere cogni-

tivo. Secondo Moro, il riemergere di quest'ultima fallacia, foriera di tragedie passate non così lontane dal nostro presente, dipenderebbe in linea diretta dall'interpretazione in chiave funzionale delle differenze linguistiche, e questo perché, «dal momento che non può essere negato il rapporto tra pensiero e linguaggio, [...] una graduatoria tra lingue può fornire il pretesto, se non la ragione, di una graduatoria tra intelligenze» (p. 21), sovrapponendo così la questione della superiorità razziale a quella della superiorità linguistica.

Se nel primo capitolo dell'opera Moro mette in guardia il lettore dall'associare le differenze formali rilevabili tra le lingue ad un loro diverso modo di esprimere più o meno correttamente i contenuti del pensiero, nel secondo egli si propone di individuare i fattori fondamentali che determinarono l'emergere di quest'idea in linguistica. Come rileva l'autore, se l'idea che «la lingua sia lo specchio della mente» (p. 29) era presente già nel Seicento nella concezione del linguaggio dei grammatici di Port-Royal, l'idea di una differenza nel modo in cui le lingue riproducono il pensiero emergerà soltanto nell'Ottocento ad opera di Friedrich Schlegel e Wilhelm von Humboldt, i quali sancirono il primato del tipo linguistico rappresentato dalle lingue indoeuropee e del quale il sanscrito era considerato l'esponente più illustre, ossia quello flessivo. In particolare, fu Schlegel a tracciare una bipartizione tra le lingue flessive e tutte e altre: le prime erano da lui considerate "organiche" e vitali, e per questo maggiormente adatte a manifestare il pensiero; le altre, definite come "meccaniche", erano invece descritte come arbitrarie, fragili e imperfette. Negli anni Cinquanta dell'Ottocento, continua l'autore, un altro fattore che contribuì in modo determinante al rafforzarsi del rapporto tra lingua e razza fu il successo delle teorie linguistiche di Max Müller, il quale postulò un legame diretto tra i tipi isolante, agglutinante e flessivo e lo sviluppo sociale delle rispettive popolazioni, con il corollario che il passaggio dal nomadismo ad una società basata sullo Stato fosse appannaggio solamente delle popolazioni associate alle lingue del terzo tipo, ossia quelle indoeuropee, da Müller denominate "ariane". L'epilogo di tutto questo percorso ideologico, lamenta Moro, si ebbe infine con l'approdo, nella seconda metà dell'Ottocento, alla posizione sostenuta da Honoré Joseph Chavée, il quale arrivò a sostenere l'idea che «a due lingue diverse corrispondono necessariamente anche due organizzazioni cerebrali diverse» (p. 45). A conferma del

quadro già delineato da Moro si potrebbe aggiungere come il razzismo linguistico divenne esplicito nelle riflessioni di Frederic Farrar, per il quale uno dei meriti della scoperta del sanscrito fu quello di svelare l'unità della grande razza indoeuropea o ariana, la quale avrebbe guidato verso il progresso «le altre razze, siano esse uguali o inferiori» (*Families of Speech*, Londra 1870, p. 187, mia traduzione), e in quelle di James Byrne, secondo il quale ciò che starebbe alla base dell'aspetto peculiare delle lingue indoeuropee e di quelle semitiche, e cioè della flessione, non sarebbe altro che «una superiorità di potere mentale» (*General Principles of the Structure of Language, Volume II*, Londra 1885, p. 275, mia traduzione) delle rispettive razze.

Nel terzo capitolo l'autore offre quella che ritiene essere la soluzione al problema del razzismo linguistico, o meglio, la attribuisce a Noam Chomsky, la cui teoria innatista «radica il linguaggio nella struttura biologica degli esseri umani in un modo che non si era mai visto prima» (p. 49). Prima di discutere i fondamenti della visione chomskiana del linguaggio, delle lingue e delle loro differenze, Moro percorre le tappe dello sviluppo della linguistica del Novecento partendo dall'inizio del secolo e, più precisamente, dalla figura di Ferdinand de Saussure. L'importanza di Saussure risiede, secondo l'autore, nell'aver teorizzato per primo la natura sistemica del linguaggio e l'interrelazione dei suoi elementi costitutivi. Questi due concetti, riassumibili nell'«idea [...] che il valore di un elemento in una lingua non sia mai assoluto ma dipenda dal valore degli altri» (p. 53), nella prima metà del Novecento caratterizzeranno anche il pensiero dei linguisti strutturalisti americani i quali, come Saussure, consideravano la lingua non un sistema cognitivo ma un sistema di elementi. Ed è proprio tenendo conto di questo quadro, nel quale «lo studio delle lingue si era per il momento tenuto lontano dalle sabbie mobili della questione del rapporto tra strutture linguistiche e strutture neurobiologiche» (p. 58) che, secondo l'autore, devono essere valutate l'originalità e la portata della teoria sintattica formulata da Chomsky. In particolare, la rivoluzione di Chomsky e della sua teoria, la cosiddetta “grammatica generativa”, secondo Moro «scaturì dalla comprensione che tre fenomeni apparentemente *indipendenti* si spiegavano come effetti di *un'unica causa*» (pp. 64-65): primo, che la competenza linguistica del parlante è basata su regole la cui astrattezza e complessità non ne permettono l'acquisizione tramite un insegnamento

esplicito; secondo, che le relazioni tra le parole non sono di tipo *lineare*, ma *gerarchico*; terzo, che «i bambini imparano spontaneamente qualsiasi lingua» non soltanto in un tempo sorprendentemente rapido considerato il livello di complessità di tale compito, ma anche «in un tempo medio praticamente identico» (p. 65). Fu quest'ultima considerazione, ossia la disparità tra lo stimolo linguistico ricevuto dai bambini nella loro prima infanzia e la complessità delle grammatiche da essi acquisibili, che suggerì a Chomsky che alla base della capacità di linguaggio vi fosse «un'impalcatura [...] comune a tutte le lingue» (p. 66) innata in ogni essere umano e, pertanto, geneticamente determinata. L'esposizione di Moro non si limita alla sola analisi della teoria chomskiana, ma intende soprattutto sottolinearne sia l'importanza in chiave "antirazzista", sia la valenza empirica. Riguardo alla prima, l'autore insiste sulla portata innovativa della grammatica generativa rispetto al contesto culturale degli anni Cinquanta del Novecento. In particolare, se da un lato la maggior parte dei linguisti dell'epoca considerava la dimensione della variazione linguistica come potenzialmente illimitata, dall'altro Chomsky poneva come presupposto essenziale della capacità umana di acquisire un linguaggio proprio l'esistenza di «un certo numero limitato di gradi di libertà» (p. 66), ossia dei limiti innati che potessero effettivamente garantire al bambino la possibilità di convergere le proprie ipotesi di apprendimento su una data grammatica in un lasso di tempo sorprendentemente breve. Ed è proprio questa, per Moro, la soluzione al problema del razzismo linguistico. Secondo l'autore, infatti, il fatto che «tutti gli esseri umani [nascono] esattamente con le stesse istruzioni geneticamente determinate per arrivare a costruire e interpretare *tutte* le lingue possibili» (p. 70) implica necessariamente che le lingue non sono altro che variazioni su un unico tema, e la questione dell'esistenza di lingue migliori o peggiori di altre cessa pertanto di porsi. Sebbene l'esistenza di una tale impalcatura innata possa sembrare poco plausibile, Moro evidenzia come un simile principio di ridondanza sia alla base del funzionamento del nostro sistema immunitario, come ipotizzato dal premio Nobel per la medicina Niels Jerne quando suggerì che «almeno una parte del repertorio degli anticorpi sia parte della nostra dotazione genetica, ben prima di ogni esperienza» (pp. 67-68). Un ulteriore supporto alle teorie generativiste, ed in particolare all'idea che il linguaggio umano costituisca un fatto

unico tra gli esseri viventi, sarebbe fornito secondo l'autore da alcuni esperimenti condotti a partire dagli anni Settanta sugli scimpanzé e sui gorilla. Pur arrivando ad apprendere un sistema complesso come la lingua dei segni americana, infatti, in nessuno dei casi esaminati tali animali si dimostrarono in grado di generare nuovi significati nei loro enunciati modificando l'ordine delle parole, come invece facciamo noi esseri umani quando assegniamo a due frasi come *Caino uccise Abele* e *Abele uccise Caino* due significati diversi. L'ultima prova presentata da Moro a favore delle ipotesi generativiste proviene da alcuni esperimenti di tipo neurolinguistico condotti da egli stesso e da altri studiosi (si veda per esempio, sempre di A. Moro, *I confini di Babele. Il cervello e l'enigma delle lingue impossibili*, Bologna 2015), e riguarda il cosiddetto "principio di dipendenza della struttura": vale a dire, il principio secondo il quale le parole non sarebbero organizzate nella frase secondo un ordine di tipo lineare, bensì in base ad un ordine di tipo gerarchico che sarebbe a propria volta il prodotto della capacità unicamente umana di poter combinare ricorsivamente tra loro coppie di parole, nonché di sintagmi, un numero potenzialmente infinito di volte. L'ipotesi sperimentale era la seguente: se, come sostiene Chomsky, la dipendenza dalla struttura rappresenta realmente una delle «proprietà molto generali che sono effettivamente comuni a tutte e solo alle lingue umane» (p. 77), qualora un apprendente si trovasse di fronte al compito di apprendere delle regole grammaticali "impossibili", cioè «regole lineari contrarie al principio di dipendenza della struttura» (p. 89), il cervello di quest'ultimo reagirebbe necessariamente in modo diverso dal modo in cui reagisce di fronte a regole possibili. Gli esperimenti verificarono tale ipotesi, rilevando nell'acquisizione delle regole impossibili l'attivazione di diverse reti neurali rispetto a quelle attivate durante l'acquisizione delle regole possibili.

Come si è occupato di dimostrare che le differenze di tipo sintattico non implicano differenze nell'elaborazione del pensiero, similmente nel quarto capitolo Moro intende dimostrare che «anche se si passa [...] alle parole singole [...] la situazione non cambia» (p. 99). Contestando la concezione di Bruno Snell secondo cui una lingua come il latino non sarebbe in grado di esprimere le stesse idee e gli stessi ragionamenti di una lingua come il greco per il fatto di non disporre, a differenza della seconda, dell'articolo determinativo,

Moro mostra come «anche a livello lessicale, oltre che sintattico, esistono per così dire delle “guide” che precedono l’esperienza [...] e per questo universali su base biologica» (p. 105). Il primo ambito affrontato è quello delle parole funzionali, in merito alle quali l’autore cita il lavoro di ricerca di John Barwise e Robin Cooper sui quantificatori come dimostrazione dell’universalità della componente logica del linguaggio. Passando al dominio delle parole non funzionali, Moro insiste particolarmente sulla pericolosità dell’idea che dalla lingua dipenda la visione della realtà, come sostenuto negli anni Trenta dalla cosiddetta teoria “Sapir-Whorf”. Essa è descritta dall’autore come «un rigurgito [...] dei tentativi ottocenteschi di ancorare lingua e cultura in modo inscindibile, vera base di ogni razzismo linguistico, poi smontata dai linguisti stessi» (p. 112) ma ancora viva nel pensiero comune. Sarebbe troppo lungo soffermarci qui sui vari esempi presentati a sostegno dell’idea che la nostra percezione della realtà non è veicolata dalla lingua che parliamo; ci limiteremo a riportare due casi nei quali «si usa la linguistica per modificare aspetti sociali ed etici» (p. 118), ossia la questione del genere grammaticale e quella del lessico legato alla disabilità. Nel primo caso, se in una lingua come l’italiano estendere l’uso del femminile laddove la lingua prevede il maschile può certamente avere l’effetto di sottolineare la disparità di genere, Moro lamenta non solo che una tale operazione comporta comunque, a livello linguistico, delle conseguenze indesiderate, come ad esempio l’aumento del livello di ridondanza qualora si scelga di affiancare sempre il femminile al maschile, ma anche e specialmente che «non si può sperare di avere alcuna garanzia di costruire società migliori agendo soltanto sulla lingua, deresponsabilizzandosi con questo da impegni concreti» (p. 123). Il parallelismo tra l’uso del termine “razza” e la questione del genere grammaticale si riscontra anche, e forse maggiormente, in quella del lessico legato alla disabilità. Secondo Moro, l’uso «politicamente corretto di negazioni [...] – non così, diversamente così – » (p. 126), oltre a rimanere comunque sgradevole per la persona cui tali termini vengono riferiti, fornisce una sorta di alibi che ci esonera dall’accettare la diversità e dal valorizzarla nella vita di tutti i giorni, non solo “sulla carta”.

Il quinto capitolo del volume è dedicato al supporto dell’idea, già introdotta nel primo capitolo, della sostanziale equivalenza della complessità delle lingue. In favore di questa tesi, Moro porta due argomenti. Il primo

è questo: se l'italiano, lingua dalla morfologia verbale relativamente ricca, fosse davvero più complessa di una lingua morfologicamente più semplice, come per esempio l'inglese, il tempo di apprendimento medio della prima sarebbe maggiore rispetto a quello della seconda. Sebbene questo possa effettivamente accadere negli adulti, per i quali tuttavia imparare una seconda lingua rappresenta «un'operazione soggettiva totalmente condizionata dalla lingua già acquisita» (p. 133), ciò non accade nei bambini, i quali apprendono qualsiasi lingua con la stessa rapidità, a prescindere dalla ricchezza morfologica o da qualsiasi altro parametro di natura grammaticale che le caratterizzi. Ad integrazione del quadro delineato da Moro si potrebbe anche citare lo studio condotto da Theresa Biberauer, Anders Holmberg, Ian Roberts e Michelle Sheehan ("Complexity in comparative syntax: the view from modern parametric theory", in *Measuring Grammatical Complexity*, Oxford 2014, pp. 103-127), il quale non solo conferma il fatto che lingue apparentemente semplici possono essere comunque caratterizzate da aspetti grammaticali alquanto complessi e rari, come per esempio il sistema dell'ausiliare in inglese, ma anche e soprattutto che, conformemente alla visione chomskiana, la complessità delle lingue è da considerarsi un fattore puramente estrinseco in quanto relativo non a delle supposte differenze nella nostra componente biologica, bensì all'organizzazione dei dati linguistici primari su cui si basa il processo di apprendimento. Il secondo argomento, Moro lo ricava dalle indagini sperimentali, le quali mostrano come le regole sintattiche delle varie lingue, a prescindere dal loro percepito grado di complessità, attivano le stesse reti neuronali. L'ultimo mito da sfatare, secondo l'autore, è l'idea che le lingue vadano incontro ad un inesorabile deterioramento. Tale tesi, sviluppata nell'Ottocento soprattutto ad opera di August Schleicher, secondo il quale la storia delle lingue successiva alla nascita dell'indoeuropeo corrispondeva ad una fase di inesorabile decadenza sia fonologica che morfologica, pur essendo stata abbandonata dai linguisti stessi rimane, secondo Moro, viva e vegeta nell'immaginario comune. Egli cita, come esempi di ciò, la connotazione negativa associata al linguaggio dei cosiddetti "nativi digitali", ricco di abbreviazioni; oppure, il progressivo abbandono, in italiano, del congiuntivo; o ancora, l'idea che le lingue debbano essere migliorate. In generale, la visione di Moro è che, diversamente da quanto suggerito dal pensiero comune, la

direzione nella quale si muovono le lingue è essenzialmente neutrale, e questo in quanto ognuna di esse è e sarà sempre in grado di esprimere qualsiasi contenuto: sia essa ricca di abbreviazioni, come lo era per esempio il latino delle epigrafi, o «anche senza distinzione tra modo condizionale e congiuntivo; almeno, così se la cavava Shakespeare in inglese» (p. 140).

L'ultimo capitolo del volume, a dispetto del titolo, non si limita solamente a tracciare un bilancio conclusivo del rapporto tra le idee del pensiero comune e il pericolo del riemergere di un razzismo linguistico, ma offre anche alcuni spunti di riflessione riguardanti l'importanza delle differenze linguistiche nello studio del linguaggio. Riguardo alla questione del razzismo linguistico, Moro insiste sulla natura subdola di alcune convinzioni condivise, tanto apparentemente innocue quanto profondamente radicate: per esempio, sostiene l'autore, non esistono lingue che aprono la mente, bensì autori in grado di aprire la mente, i quali ci hanno tramandato le loro opere proprio in quelle lingue; e se lo studio di una determinata lingua ci apre la mente più che studiarne altre, ciò dipende non da quella lingua in sé, ma dal modo in cui la si studia e dalla conoscenza che si ha di essa. Riguardo invece allo studio del linguaggio, Moro attribuisce alla variazione linguistica un'importanza fondamentale. Solo se lo studio approfondito delle singole lingue si accompagna alla loro comparazione, sostiene l'autore, la linguistica può sperare di far luce sulla natura del linguaggio. La variazione, in questo senso, appare come fattore unificante che può permetterci, passo dopo passo, di stringere il cerchio attorno alla componente innata e universale alla base di tutte le lingue e, insieme, del nostro essere umani.

La razza e la lingua. Sei lezioni sul razzismo è dunque un libro adatto sia a chi sia interessato specificamente alla linguistica e alla sua storia, sia a chi, pur non disponendo di conoscenze pregresse in questa disciplina, volesse approfondire in modo accessibile e gradevole la questione del razzismo da un punto di vista forse insolito, ma senza dubbio ricco di spunti di riflessione e più che mai attuale.

Alessandro Riolfi

Si ringrazia per la collaborazione
il dott. Ferdinando Longobardi



IL TORCOLIERE • Officine Grafico-Editoriali d'Ateneo
Università di Napoli L'Orientale
stampato nel mese di dicembre 2023

