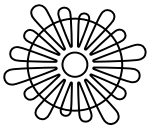
Marco ANGSTER

At the boundaries of word-formation

Contrastive coreference in 30 European languages



Sveučilište u Zadru
Universitas Studiorum
Jadertina | 1396 | 2002 |

Marco Angster

AT THE BOUNDARIES OF WORD-FORMATION

CONTRASTIVE COREFERENCE IN 30 EUROPEAN LANGUAGES

Publisher / Nakladnik
University of Zadar / Sveučilište u Zadru

For the publisher / Za nakladnika
Josip Faričić, Rector / rektor

Publishing committee /
Povjerenstvo za izdavačku djelatnost
Lena Mirošević, Chair / predsjednica

Reviewers / Recenzenti
Matea Andrea Birtić
Fabio Montermini

Language editor / Lektorica
Sandra Mlađenović

Cover photo /
Fotografija naslovnice
Mia Batinić Angster

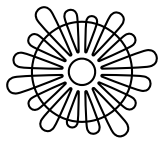
Design & Layout /
Grafičko oblikovanje i prijelom
University of Zadar / Sveučilište u Zadru

ISBN
978-953-331-442-6

Marco ANGSTER

At the boundaries of word-formation

Contrastive coreference in 30 European languages



Zadar, 2023.

CONTENTS / KAZALO

CONTENTS / KAZALO

CONTENTS / KAZALO

Acknowledgments

On the 15th February 2013, I defended my PhD thesis (Angster 2012) at the University of Pavia (Italy). The topic of the thesis – entitled Le marche di coreferenza contrastive nella formazione delle parole. Una panoramica delle lingue d'Europa – was inspired by a lecture by Ekkehard König on "reflexive compounds", held in Pavia in November 2009 as part of the curriculum of the doctoral school, directed in those years by the late Marina Chini. This was my first activity in the doctoral program, into which I was admitted after a written and an oral exam. During the latter, I defined my intention to work on word-formation from a cross-linguistic perspective. I was thus looking for a topic that involved a word-formation pattern and could be studied by comparing complex words found in different languages. König's lecture – later published as König (2011) – was crucial in the choice of topic for my dissertation. I am, however, indebted to my main supervisor, Davide Ricca, for the idea of pursuing a cross-linguistic study on word-formation, and to my other supervisors – Anna Giacalone Ramat, Nicola Grandi, Gianguido Manzelli and Nicoletta Puddu – for their help, suggestions, and support in dealing with a topic at the crossroads of several subfields of linguistics.

This book is a re-elaboration of my dissertation. More precisely, Chapters 1 and 2 of this book partially depend on small parts of Chapters 1 and 2 of my dissertation, but have been written almost completely from the ground up, as was Chapter 3. Chapters 4, 5 and 6 constitute a translation from Italian into English of Chapters 3 and 4 of the dissertation, which have, however, been revised, updated and completed in various ways, especially as far as the insights based on Construction Morphology are concerned.

The idea of publishing my dissertation as a book was seen as a priority by my supervisors ever since the first months after the defense. However, as the years passed and new commitments and projects piled up, the idea, though constantly haunting me, was becoming less and less likely, while the exact focus of a potential book oscillated between a limited comparison of Italian and German alone (i.e. Chapter 2 of the dissertation), and, alternatively, a larger cross-linguistic (Chapter 3) and diachronic survey (Chapter 4), encompassing thirty European languages included in the sample considered in the dissertation.

It took me almost ten years to feel the needed pressure to resume my project and

work on this English re-elaboration of my dissertation. I am particularly grateful to Livio Gaeta for encouraging me in this direction, and for his many precious suggestions during the past years. I would also like to thank the reviewers of the book – Matea Birtić and Fabio Montermini – for the careful reading of the manuscript and their suggestions for its further improvement. I warmly thank my colleagues at the Department of Linguistics of the University of Zadar for their support during the months spent preparing the manuscript. I am also thankful to the University of Zadar and its publishing office for their support.

A lot of friends and colleagues shared smaller or bigger portions of this ten-year journey: Raffaele Cioffi, Silvia Dal Negro, Martina Di Stefano, Maicol Formentelli, Francesca Franzon, Vincenzo Galatà, Elisa Ghia, Eugenio Goria, Frane Malenica, Chiara Meluzzi, Matteo Rivoira, Eleonora Sausa, Chiara Zanini (to mention but a few, with apologies to those who have been omitted).

The crucial help for completing this effort came, however, from my family both in the Alps and on the Adriatic Sea.

Abbreviations

The following abbreviations for grammatical morphemes and morphosyntactic categories are mainly used in the glosses to the examples and in the representation of constructional schemas.

Abbreviations referring to lexicographic works quoted in the text are listed below the references.

°	possible, but unattested
*	ungrammatical
>	diachronic derivation
	higher in a hierarchy
≈	paradigmatic relation
→	synchronic derivation
←	backformation
↔	links in the Parallel Architecture
1	first person
2	second person
3	third person
A	adjective
ABS	absolutive
ACC	accusative
AGR	agreement
AOR	aorist
ART	article
AUX	auxiliary
CC	contrastive coreference
CP	contrastive possession
DAT	dative
DEF	definite
DEM	demonstrative
DEP	deponent
EMPH	emphatic
ERG	ergative
F	feminine
FUT	future
GEN	genitive
ID	identity function
IMPR	impersonal
IND	indicative
INSTR	instrumental
INT	intensifier
INT adN	adnominal intensifier
INT adV	exclusive adverbial intensifier
INT att	attributive intensifier
INTS	intensive pronoun
LINK	linking element
LOC	locative
M	masculine
MID	middle
N	noun
NOM	nominative
NPST	nonpast
Num	numeral
P	preposition
PASS	passive
PFCT	perfect
PL	plural
POS	part of speech
POSS	possessive

PREV	preverb
PRO	pronoun
PRS	present
PRT	particle
PRTT	partitive
PST	past
PTCP	participle
Q	quantifier
ℜ	implicit relationship
REC	reciprocal
REFL	reflexive pronoun
REFL H	heavy reflexive pronoun
REFL L	light reflexive pronoun
SCV	separable complex verb
SG	singular
SOC	sociative
SV	subjective version
sW	semi-word
V	verb
VN	verbal noun
W	word

1

INTRODUCTION

This book constitutes a survey of the way in which contrastive coreference is marked in European languages. I preliminary define contrastive coreference as the meaning, conveyed by a number of word-formation elements in different languages, that can be paraphrased in relevant formations, either by using a heavy reflexive or an adverbial intensifier (see König and Gast 2006). An example of these markers and formations can be a complex word such as Italian *autodistruzione,* 'self-destruction' in (1), whose base, *distruzione* 'destruction', is an action noun derived from the verb *distruggere* 'destroy', while the first element, *auto-* – corresponding to English *self-* – indicates that the action of destroying is undergone by the same referent who is performing the action, and can be paraphrased by using a reflexive sentence:

(1) *la mia* ***autodistruzione***
'my self-destruction'
(2) *Io* ***distruggo me stesso***
'I destroy myself'

Less often, if a referent, especially inanimate, affected by the action of destroying is made explicit, the same formation can be paraphrased with a sentence containing an adverbial intensifier:

(3) *il progetto di* ***autodistruzione*** *del suo stesso mito* [itTenTen16][1]
'the project of self-destruction of his own myth'
(4) ***Da sé distrugge*** *il proprio mito.*
'He destroys his own myth by himself.'

A marker like Italian *auto-* constitutes a compelling object of study because it operates on the morphological level, forming new lexemes, but it also contributes semantically, in a way that corresponds to highly grammatical meanings expressed by quite diverse syntactic constructions.

Conducting a cross-linguistic survey of contrastive coreference markers, such as Italian *auto-,* is a challenge for morphologists, because formations very similar

[1] The corpus itTenTen16 is part of the TenTen family of corpora (Jakubíček *et al.* 2013), accessible via the SketchEngine platform (Kilgarriff *et al.* 2004). The corpus deTenTen13, used to extract some German examples for the book, belongs to the same family.

to the ones found in Italian are available in many European languages, but their formal features, such as the lexical sources of the relevant markers and the classification of these processes in different languages may vary greatly, ranging between derivation and compounding, and sometimes bordering on syntactic structures.

Additionally, the core domain of productivity of all contrastive coreference markers is that of deverbal action nominals, while – quite paradoxically for a device which apparently operates on the argument structure of verbs – their combination with verbal bases, especially in predicative function and in finite verb forms, is rather marginal in many European languages, if available at all.

In his pilot study on what he calls "reflexive (nominal) compounds", König (2011: 125-126) points out that these formations «are highly condensed expressions of very complex meanings and seem to be found primarily in languages with a long tradition of literacy, philosophical writing and psychological reflections», and «[m]ore often than not the relevant pattern seems to have been borrowed from neighboring languages and Greek seems to have played an important role as a source». This suggests that all these formations might in some way be related in different languages, and it raises the question whether the relevant language-specific markers are productive, or if their presence in different languages depends exclusively on item-based, imitative neologisms that have emerged after single loan formations had been diffused through cultural contacts.

In this book, we will expand on the work started by König (2011), and give an account of contrastive coreference markers in the languages of Europe from both a synchronic and diachronic point of view. In the next chapter, we will survey theoretical issues implied by the study of contrastive coreference markers, while in the third chapter we will present the language sample and the methodological issues linked to the cross-linguistic study of wordformation.

The fourth and fifth chapter will include a synchronic survey of contrastive coreference markers (henceforth $_{CC}$markers) in European languages. In the fourth chapter, we will consider $_{CC}$markers from the point of view of their formal make-up, in an attempt to define whether individual markers found in specific languages belong to derivation or compounding. In the fifth chapter, we will shift our attention to the bases with which $_{CC}$markers combine, and we will put forward some possible generalizations and tendencies observable in the data.

With the sixth chapter, the perspective will shift from synchronic to diachronic. The $_{CC}$markers identified and the tendencies observed will be accounted for in

terms of linguistic and extra-linguistic – mainly cultural and historical – aspects.

The seventh and final chapter will constitute a summary and critical assessment of the results of this study, which will also suggest some possible future directions in the cross-linguistic study of contrastive coreference and word-formation phenomena.

1.1 Intensifiers and reflexives in complex words

The existence of complex words involving deverbal nouns modified by markers that operate on the argument structure of the underlying verb leading to a reflexive interpretation – we will provisionally call them "reflexive word-formation markers" – has been referenced in the literature in English and German at least since Edmondson and Plank (1978). Since then, several contributions appeared, especially those dealing with individual languages, such as English (Di Sciullo 1996), Italian (Mutz 2003, 2009, Castella 2010), French (Dugas 1992, Mutz 2004), Spanish (Arquiola 2003, Orqueda *et al.* 2020), Greek (Alexiadou 2015), and a few studies involved the discussion of data from small cross-linguistic language samples, along the lines of König's (2011) pilot study (Angster 2012, Vasilova 2017).

These contributions range from descriptive to more theoretical studies, investigating – e.g. Arquiola (2003) – how the study of complex words involving a reflexive interpretation can shed light on the division of labor between different levels of description (e.g. syntax, morphology and lexical semantics) in the construction of the meaning of complex words, starting from their building blocks, i.e. the base verbs and the relevant "reflexive word-formation markers".

With regards to theoretical approaches in this field, Alexiadou's (2015) is linked to distributed morphology (Halle and Marantz 1993), and is essentially based on intuition: the paper focuses on the role of the meaning of roots in influencing the morpho-syntactic combinations which they can enter. Other contributions, such as Orqueda *et al.* (2020), are grounded in Construction Grammar (Goldberg 2006): the focus of the article is on the syntactic use of verbs prefixed with *auto-* in Spanish, rather than on the function of the prefix in word-formation in general.

The aforementioned studies also differ when it comes to the focus on subsets of formations. Some studies consider mainly deverbal action nouns modified by

"reflexive word-formation markers" (Edmondson and Plank 1978, König 2011), while others focus on complex verbs (Dugas 1992, Arquiola 2003, Alexiadou 2015, Orqueda *et al.* 2020). None of these studies concentrate on the case of deverbal adjectives occurring in combination with "reflexive word-formation markers", even though they often mention them. The difference in the focus of individual studies also depends on the languages considered, and on the difference in productivity of the relevant markers with bases belonging to the classes of (deverbal) nouns and adjectives and to that of verbs. Nominals appear as the most productive domain for different "reflexive word-formation markers", followed by adjectival and verbal bases, the latter ones being almost completely excluded from combining with *self-* and *selbst-,* in English and German respectively (König 2011: 125).

A common discussion found in the literature concerns the constraints on the verbal bases that can combine with "reflexive word-formation markers". Generally, it is observed that transitive, other-directed verbs are the main domain of productivity (König 2011: 122; Alexiadou 2015: 63). This excludes intransitive – more specifically, mono-argumental – verbs, and typically self-directed verbs such as body care/grooming verbs, i.e. verbs belonging to the domain of the Middle (s. Kemmer 1993). This parallels with the constraints to which verbs occurring in emphatic reflexive constructions are subject. For Greek *afto-,* Alexiadou (2015: 66) adds a constraint concerning the passivizability of the base verb, by which she also precludes transitives forming anti-causatives from occurring with *afto-*. In addition, she shows that only agentive/causative accomplishments are compatible with *afto-* (Alexiadou 2015: 69). However, these constraints are subject to cross-linguistic variation: the incompatibility of Greek (but also of Romance languages) with subject-experiencer predicates such as 'love', 'hate' – even if bi-argumental and passivizable – is not shared by languages like English (*self-love*, *self-hate*) or German (*Selbstliebe*, *Selbsthass*). Also, the incompatibility of verbs in the anti-causative alternation is far from absolute: Mutz shows how *auto-* productively combines with anti-causative verbs in Italian (*l'autoaccensione del fieno* 'the self-kindling of the hay') and French (*l'autoinflammation du gaz* 'the self-ignition of the gas' Mutz 2004: 357).

Languages can also differ as far as the number of competing "reflexive word-formation markers" is concerned. König (2011: 123) points out how in German three subsets of "reflexive compounds" can be found, one characterized by

the borrowed prefix *auto-* (5).a, the other two including formations composed of the intensifiers *selbst* (5).b and *eigen* (5).c.

(5) a. *Autodidakt* 'self-educated person', *Autokratie* 'autocracy', *Autobiographie* 'autobiography';
b. *Selbstbetrug* 'self-deception', *Selbstverschulden* 'one's own fault', *Selbstlob* 'self-praise';
c. *Eigenleistung* 'personal contribution', *Eigenwert* 'intrinsic value', *Eigenmittel* 'one's own resources', *Eigenmächtigkeit* 'unauthorized behavior'.

The competition of different markers is not exclusive to German, and it is recognized by König (2011) as one of the possible parameters of cross-linguistic variation of "reflexive compounds". He mentions the case of Finnish (*itse-* vs. *oma-*), Modern Greek (*afto-* vs. *iðio-*) as well as English (*self-* and the marginally productive *auto-*) as examples of competition, and of Chinese (*zì*) as an example of the presence of a single marker (König 2011: 125). Further examples of competition are observed in Vasil'ová (2017) for Hungarian (*ön-* vs. *maga-*), Polish (*auto-* vs. *samo-*) and Slovak (*samo-* vs. *seba-*/*sebe-*). On the other hand, she ascribes Italian, Romanian, English and Dutch to the languages where only one marker is used in what she calls, following König (2011), "reflexive nominal compounds".

To my knowledge, there is no example in the literature of any attempt to compare languages where more than one marker is available to identify different patterns of competition between different strategies. König (2011: 123) suggests that the competition of *selbst-* and *eigen-* in German is characterized by both overlapping areas (*Selbstlob*/*Eigenlob* 'self-praise') and specialization: e.g. *eigen-* appears to be compatible with non-derived bases (*Eigenblut* 'one's own blood'), while a «vague possessive relation (*appartenance*) is still apparent in compounds with *eigen-*, and more clearly so in the possible paraphrases with attributive *eigen*: (*sein eigener Name — sein Eigenname*; *sein Eigenheim — sein eigenes Heim* 'his own home'; [...])». On the other hand, the «adverbial exclusive type is rare among compounds with *eigen*, the only exception being adjectival compounds like *eigenhändig* < *mit eigenen Händen* 'with one's own hands'), *eigenmächtig* (< *aus eigener Macht*) 'high-handed', which can themselves be used as exclusive adverbial intensifiers» (König 2011: 123). He also suggests that the competing

strategies of Finnish are parallel to those found in German and Modern Greek, but he does not provide any further detail of this alleged parallel.

1.2 Functions of "reflexive word-formation markers"

This leads us to the topic of the meaning or function of "reflexive word-formation markers". Beyond the rather wide variety of approaches and interpretations of the contribution of such markers in the relevant complex words, an aspect that is generally common to all scholars who have considered the problem is that these markers – either when defined as prefixes or when recognized as compound elements – convey a rather grammatical meaning, corresponding to that of reflexive pronouns or constructions, of various kinds of intensifiers, or even of argument alternations, such as anticausative verbs. The functions recognized are thus connected mainly to valency changes, but also to pragmatic operations, such as contrastive focus.

I will now focus on the functions suggested by Mutz (2003, 2004, 2009) for Italian and French *auto-,* and by König (2011) mainly for German *selbst-*.

In a series of studies focusing on Italian (Mutz 2003, 2009) and French (Mutz 2004) lexicographical data, Katrin Mutz deals with complex formations containing the prefix *auto-*.[2] As a starting point, Mutz compares the functions of the reflexive pronoun *si* with those of *auto-*. Mutz (2009) lists the functions of Italian *si* based on Monachesi (1999): the list with the relevant examples provided by Mutz is presented under (6) (Mutz 2009: 310-311; my translation, emphasis of the author).[3]

(6) 1. *si* is a *reflexive* pronoun, e.g. *Anna si difende* 'Anna defends herself' (vs. *Anna difende Pietro* 'Anna defends Pietro');
2. *si* is a *reciprocal* pronoun, e.g. *Anna e Gianni si baciano* 'Anna and Gianni are kissing (each other' (i.e. 'Anna kisses Gianni and Gianni kisses Anna');

[2] Mutz derived the Italian data from dictionaries like disc and Palazzi-Folena, while the French data were derived from the dictionary Nouveau Petit Robert. Additional data were collected online in a Google search carried out in 2001.

[3] Mutz (2003) is originally written in Italian. The translation of the original text and of the Italian examples is mine.

3. *si* is an *anticausative* element, e.g. *la porta si apre* 'the door opens' (vs. *Anna apre la porta* 'Anna opens the door');
4. *si* highlights the theme or patient of an action (so called *si passivante*), e.g. *i libri si vendono bene* 'books sell well';
5. *si* signals the genericity of the agent of an action (so called *si impersonale*), e.g. *in questa città si va spesso al cinema* 'in this town they often go to the cinema';
6. *si* conveys the *inchoative* aspect, e.g. *addormentarsi* 'to fall asleep', *arrabbiarsi* 'to get angry';
7. *si* is found in a number of lexicalizations (without adding any specific semantic contribution), e.g. *comportarsi* 'to behave', *vergognarsi* 'to be ashamed'.

According to Mutz, in complex words with *auto-,* the functions of *si* reproduced by the prefix are exclusively those set out in points 1. and 3. in (6) above. With regard to the other uses listed, Mutz shows how *auto-* does not indicate reciprocity (cfr. point 2.: **autobaciamento di Anna e Gianni* 'self-kissing of Anna and Gianni'), has no passive value (cfr. point 4.) in *l'autovendita dei libri* 'self-sale of books', nor are there suitable examples of an impersonal value (cfr. point 5.; Mutz 2003: 653).[4]

According to Mutz, these two functions operate on the argument structure of the base verb. The reflexive function would indicate the coreference of the two arguments – agent and patient in *autopunizione* 'self-punishment', agent and goal in *autoriferimento* 'self-reference' – of a transitive (or more generally bi-argumental) verb (7).

(7) (Mutz 2003: 653)
a. *la* punizione *di Gianni* 'the punishment of Gianni'
b. l'autopunizione *di Gianni* 'the self-punishment of Gianni'
c. **l'autopunizione di Gianni da parte di Anna* 'the self-punishment of Gianni by Anna'

[4] It can be added that the marker *si* is labile in derivation, so that in the (attested or possible) nominalizations corresponding to the verbs under points 6. and 7., only the theme of the verb (root plus possibly the thematic vowel) is preserved: *addormentarsi* → °*addormentamento*, *arrabbiarsi* → *arrabbiatura*, *comportarsi* → *comportamento*.

In (7).a, Gianni could be both the agent and the patient of the act of punishing, while in (7).b, where the "reflexive" *auto-* indicates the coreference between agent and patient, the only referent Gianni has both semantic roles. On the other hand, (7).c is ungrammatical. In essence, the reflexive function «marks a reduction of the participants in the action (compared to the action described by the base noun/verb) at the referential level and at the syntactic level of the argument structure; the semantic level, however, remains intact, the thematic roles are not eliminated» (Mutz 2003: 654; my translation).[5]

Also, in the case of the anticausative function, according to Mutz, there is only one extra-linguistic referent that assumes a semantic role in the argumentative structure. In the case of "anticausative" *auto-*, however, the role played by this referent is that of theme or patient, while the role of agent is missing. Mutz believes that the "anticausative" *auto-* has the same functions as the anticausative *si* and, following Grimshaw (1982, 1990), she believes that it operates on the argumentative structure of the verb by eliminating a syntactic argument (the external argument) and a semantic argument (the agent).

(8) (Mutz 2003: 657)
a. *l'accensione del fieno da parte di Gianni* 'the ignition of the hay by Gianni'
b. *l'autoaccensione del fieno* 'the ignition of the hay'
c. *l'autoaccensione *del fieno da parte di Gianni* 'the self-ignition of the hay by Gianni'

This leads to the anticausative argument configuration of (8).a, according to which the internal/patient argument of the transitive configuration becomes the external argument, while maintaining the semantic role of patient. The external/agent argument is eliminated, producing the ungrammatical (8).c. Mutz also notes that both the "reflexive" and the anticausative *auto-* require «as a morphosyntactic condition that the underlying base verb is at least biargumental so that, as we have seen, the reduction of the argument structure can take place» (Mutz 2003: 658).[6]

[5] «[M]arca una riduzione dei partecipanti all'azione (rispetto all'azione descritta dal nome/verbo di base) a livello referenziale e a livello sintattico della struttura argomentale; il livello semantico, però, rimane intatto, i ruoli tematici non vengono eliminati.» (Mutz 2003: 654)

[6] «[Richiedono] come condizione morfosintattica che il verbo sottostante alla base sia almeno bivalente cosicché, come abbiamo visto, la riduzione della struttura argomentale possa aver luogo» (Mutz 2003: 658)

A third function of *auto-* that Mutz identifies does not have a parallel with *si*, nor does it affect the argument structure of the underlying verb in terms of coreference or reduction of the arguments. This function, which Mutz calls "focalizing (*auto-*)" (auto- *focalizzante*; Mutz 2003: 658-61), or "excluding focalizing (*auto-*)" (auto- *escludente focalizzante)*; Mutz 2009: 313), operates at the referential and informational level. More precisely, formations with the "focalizing *auto-*" «are action nouns that put the focus on the agent of the action expressed by the base, excluding other potential agents that are more likely to act as agents. *Auto-* in this type of derivative emphatically indicates that the agent or the causer of the action is *exactly that one*, i.e., it is not the person who typically carries out the action – in the prototypical referential situation for the action described by the base». (Mutz 2009: 313; my translation, emphasis of the author).[7]

This information structure component of the "focalizing *auto-*" is also recognized as part of the other two functions. In fact, beyond differences in the manipulation of the external argument/agent, there is a function common to the three listed above, and it consists in placing the referent of the external argument in a contrastive relationship with alternative referents. «*Auto-* therefore evokes the potential existence of other agents (which are more typical for the action), and at the same time denies their actual presence». (Mutz 2009: 314; my translation).[8] This focalization on the actual agent in contrast to another expected agent is expressed differently in each of the three functions:

(9) – as for the "reflexive" *auto-*, it corresponds to the fact that no one other than the agent – co-referent with the patient/goal – participates in the action;
– as for the "anticausative" *auto-*, it corresponds to the fact that no one is the agent of the action;
– as for the "focalising" *auto-*, it corresponds to the fact that the agent differs from the expected agent.

[7] «Le formazioni sono nomi d'azione che focalizzano sull'agente dell'azione (espressa dalla base), escludendo altri agenti potenziali che siano più probabili come agenti. *Auto-* in questo tipo di derivati indica enfaticamente che l'agente o il causatore dell'azione *non sia un altro*, cioè non sia la persona che tipicamente esegue l'azione – nella situazione referenziale prototipica per l'azione descritta dalla base.» (Mutz 2009: 313)

[8] «*Auto-*, dunque, evoca l'esistenza potenziale di altri agenti (che siano più tipici per l'azione) e al tempo stesso nega la loro attualizzazione». (Mutz 2009: 314)

Despite these observations, Mutz concludes that, due to the observed polysemy, Italian *auto-* should be treated not as one, but as three different homonymous prefixes, each connected to the three functions listed in (9).[9]

As already mentioned above, König (2011) considers what he calls "reflexive nominal compounds". The author uses this term to refer to the formations, observed in some languages (English, German, French, Mandarin Chinese, Russian, Greek, Hungarian and Finnish in his sample), in which a deverbal noun combines with an element that König notes to be frequently connected to the intensifiers. Given this frequent formal coincidence that affects many European languages, but not Romance languages, König – unlike Mutz – does not compare the function of these elements with the functions of reflexivity markers such as Italian *si* or German *sich*. Instead, he considers the functions which characterize the class of intensifiers (König and Gast 2006; see 4.1).

At least four different uses have been identified in the class of intensifiers:

(10) (König 2011: 117)
a. (adnominal) *The Pope himself will come to the rally. (In contrast to one of his clergy)*
b. (adverbial, exclusive) *Mrs. Dalloway wanted to buy the flowers herself. (alone, no delegation, no help)*
c. (adverbial, inclusive) *I cannot give you any money. I am a little short of cash myself. (me too, I am not only the addressee of your request, but also in the same situation as you)*
d. (attributive) *John's own death surprised me more than that of his brother. (In contrast to s.o. else's death).*[10]

König notes that the common denominator of these uses is the contrast between a given value and alternatives relevant to it. However, this feature common to the functions of the intensifiers is not sufficient to allow all these functions to be exploited in the compounds considered by König. The functions that according to König occur in his "reflexive nominal compounds" are the adnominal, the ex-

[9] In Mutz (2004: 315-16 and notes 10-11), the proposal for three homonymous prefixes is compared to a less strong suggestion to make *auto-* a lexical entry with three different "sub-entries".

[10] In König and Gast (2006: 224), an example of the attributive use is: *Mind your own business!*.

clusive adverbial and the attributive function. The adverbial inclusive function is therefore excluded. He notes, however, that the attributive function correlates in many languages with a different marker both in syntax and in compounds – in German, for example, the contrast between *selbst* and *eigen* in syntax is apparently reproduced in their bound use in compounds. Based on this observation, he continues his discussion, considering first and foremost the adnominal and adverbial (exclusive) functions in the so-called reflexive compounds.

It should be nonetheless further specified that the adverbial exclusive function, exemplified in (10).b – which essentially corresponds to the "anticausative" and (at least partly) to the "focalizing" functions proposed by Mutz – can easily be translated into a compound, in particular in agent derivatives such as *self-provider* or *Selbstabholer* 'person who autonomously fetches certain items (mail, etc.)'. As for the adnominal function, it cannot be immediately transposed into a compound in the form expressed in (10).a, and must instead be interpreted in a more restrictive way. In the example proposed, the element under focus is the subject of the action, but it is also possible to focus on the elements that have a different syntactic function, such as the internal argument of the verb.

(11) (König and Gast 2006: 229-30)
a. *I prefer* [*the surroundings of*]F *London to London* [*itself*]$_F$.
b. [[*the surroundings of London*]] = (surroundings.of[[*London*]])
c. [[London *itself*]] = (id[[*London*]]) = [[*London*]]

In (11), we have a contrast between the surroundings of London and London, which corresponds to the contrast between the function SURROUNDINGS.OF and the identity function id that identifies London with London itself. König proposes to interpret the case of a reflexive sentence where the reflexive pronoun is strengthened by an intensifier as an application of the identity function to the reflexive pronoun. This strengthening is observed only when a language distinguishes between intensifiers and reflexives, as can be seen in (12), in the contrast between English and German.

(12) (König 2011: 117)
a. *John assesses HIMSELF*.
b. *Johann* bewertet *sich selbst*.

As an adnominal intensifier, *selbst* contrasts the object – which is coreferent with the subject of the sentence – with the possible alternative objects, which in the case of a reflexive sentence correspond to the typical objects of the action. Continuing his argument, König proposes syntactic criteria to distinguish the adnominal use and the adverbial exclusive use applied to reflexive sentences in German, showing how in the case of adnominal use, the sequence of the reflexive pronoun and of the intensifier cannot be interrupted with an adverb (13).a or with the negation (13).a, while this is possible when the adverbial function is at play (14).a-b.

(13) (König 2011: 119)
a. *Karl kritisiert (oft)* [*sich selbst*].
b. *Karl kritisiert nicht sich selbst*.

(14) (ibidem)
a. *Dieses Dorf* [*verwaltet sich (normalerweise)*] *selbst*.
b. *Dieses Dorf verwaltet sich nicht selbst*.

The two examples proposed by König – which may correspond respectively to the compounds *Selbstkritik* 'self-critic' and *Selbstverwaltung* 'self-administration' – stress the fact that one of the two semantic roles that characterize the verb arguments constitutes an unexpected choice (see also similar considerations by Mutz above). In the case of the adnominal function, this corresponds to an unexpected patient (*remarkable patient*), while in the case of the adverbial function, it corresponds to an unexpected agent (*remarkable agent*).

On the basis of these two functions of intensifiers, König proposes and exemplifies by German compounds with *selbst-* two subtypes of reflexive compounds: the adnominal reflexive compounds (whose paraphrase can be understood as 'the fact of V-ing oneself') and the adverbial reflexive compounds ('the fact of V-ing by itself').

(15) adnominal reflexive compounds (König 2011: 120)
Selbstgefälligkeit 'complacency', *Selbstzufriedenheit* 'fr. *autosatisfaction*', *Selbstironie* 'self-irony, self-mockery', *Selbstinszenierung* 'self-fashioning, self-styling', *Selbstüberschätzung* 'overestimate of

one's potential', *Selbstgespräch* 'talking to oneself', *Selbstvertrauen* 'self-confidence', *Selbstbefriedigung* 'masturbation', *Selbsthass* 'self-hate', *Selbstbild* 'self-image', *Selbstreflexionen* 'reflections', *Selbstverständnis* 'self-image', etc.

(16) adverbial reflexive compounds (ibidem)
Selbstzensur 'self-censorship', *Selbstbezichtigung* 'self-accusation', *Selbstbestimmung* 'self-determination', *Selbsthilfe* 'self-help', *Selbstachtung* 'self-respect', *Selbstentmündigung* 'self-incapacitation', *Selbstentleibung* 'suicide', *Selbstverpflichtung* 'voluntary acceptance of an obligation', *Selbstaufgabe* 'self-abandonment', *Selbstbedienung* 'self-service', etc.

As König himself notes, the two classes do not present differences from a formal point of view: in both groups, *selbst-* is always the element that acts as a modifier in the compound. Despite this, the author argues the usefulness and reality of the distinction based on the observation that in the examples in (15), the object of the action is recognizable as the remarkable participant in the coreference relationship established by reflexivity, while in those in (16), the participant who is recognized as remarkable is the subject.

König applies the distinction between adnominal and adverbial reflexive compounds also to the compounds which belong to the syntactic category of adjectives:[11]

(17) (König 2011: 121)
a. *a self-addressed envelope* → 'an envelope addressed to oneself' (adnominal → remarkable target)
b. *a self-inflicted* wound → 'a wound inflicted to his-/herself' (adverbial → remarkable agent)

As for the restrictions over the verbal bases that can access this type of structures, König observes – similarly to Mutz – that the verbs embedded in reflexive

[11] The examples of adjectives are taken from English: *self-sufficient*, *self-reliant*, *self-respecting*, *self-addressed*, *self-satisfied*, *self-deprecating* (adnominal adjectival compounds), and *self-loading*, *self-induced*, *self-appointed*, *self-inflicted*, *self-supporting*, *self-cleaning* (adverbial adjectival compounds, König 2011: 121).

compounds must be transitive, or at least have two argument positions characterized by the [+ human] trait, and they must also be other-directed (see also the proposals by Alexiadou 2015 for Modern Greek *afto-* above). The compounds in question indicate in fact that the reflexivity of the action constitutes a remarkable fact, which excludes all verbs that express a typically self-directed action, such as the verbs denoting body care (König 2011: 122, Mutz 2009: 314).[12]

1.3 Conclusion: "contrastive coreference"

It will be noted that Mutz's and König's proposals overlap in some points. First of all, both consider that the semantic contribution of *auto-* or *selbst-* to the formations in which these morphemes appear is not unitary, and different functions can be recognized. Both believe that the bases using a transitive (or at least bivalent) verb in their creation are privileged in accessing these formations, while intransitive or self-directed verbs are marginal or excluded from this type of formations. Putting aside the discussion about whether a reduction of the argument structure or of the number of extra-linguistic referents occurs in these formations, both Mutz and König note that beyond this reduction there is also a pragmatic component, characterized by a contrast between a typical configuration of the action and the configuration expressed by the formations.

The scope of their studies, however, is limited by the fact that they take into consideration almost exclusively action nouns, and in a limited way complex verbs (such as *self-destruct* or French *s'auto-féliciter* 'congratulate oneself'), overlooking other types of derivatives, such as agent nouns or deverbal adjectives – except for the mere mention of their existence and the suggestion that they share a classification similar to that of nominal derivatives (see deverbal adjectives above).

There is another aspect that is common to the proposals by Mutz and König, and perhaps limits their approach to the problem. Both consider the different functions

[12] See Haspelmath (2008) for a proposal of a frequentist explanation of asymmetries in the marking of typically self-directed actions and self-directed configurations of typically other-directed actions. Haspelmath suggests that the use of shorter markers with self-directed verbs follows from their higher frequency in self-directed usage, and supports this hypothesis with corpus data. Conversely, verbs which occur more frequently in other-directed usage are characterized by longer reflexive markers.

of *auto-* and *selbst-* exclusively as the result of these morphemes, analyzing from a limited perspective the way in which verbal bases that belong to different classes of verbs, or are characterized by different argument structures and semantic features, can affect the interpretation of *auto-* and *selbst-* by modulating their semantic contribution in the resulting formations. Furthermore, they only occasionally evaluate whether certain word-formation patterns correlate with specific interpretations of the markers.[13]

We will not deal with these issues in more detail here – for a discussion and several proposals in this regard, see Angster (2012: 73-85). What is relevant in this context is to note that, as far as the functions of "reflexive word-formation markers" are concerned, their contribution oscillates between clear cases in which they indicate that a relation of coreference holds between the arguments of a verb, and clear cases in which the thematic role of a referent in an action is highlighted as unexpected and remarkable, also implying that this is in contrast with more usual cases, where alternative referents hold the role in that particular action. The functions represented by these two opposites do not exclude each other, and in most cases co-occur in individual formations, suggesting that neither of the two can be reduced to the other. In this sense, it is curious that König (2011) defines these complex formations as reflexive compounds, despite the fact that he shows how intensifiers rather than reflexive markers are exploited in "reflexive compounds", and the subtypes that he identifies are based on the functions of intensifiers rather than on different functions of reflexive markers (as in Mutz 2003, 2004, 2009).

The function corresponding to morphemes such as *auto-* and *selbst-* is to highlight the center, and at the same time exclude everything that does not belong to it. This function also generally corresponds to the ability to establish a broad coreference relation – constituting the aforementioned center – which can involve arguments or non-arguments. In a prototypical zone, the function combines coreference and contrast with alternative (and often more expected) participants who are excluded through coreference – *auto·distruzione*/*Selbst·zerstörung* 'self-destruc-

[13] The relevance of these factors is partly raised in Mutz, with regard to the distinction between the reflexive and the anticausative functions and the close relationship between the focalizing function and verbal uses. König, on the other hand, notes how the adverbial exclusive function – in addition to being less rigidly limited to transitive verbs – is particularly relevant in agent nominal derivatives (see above).

tion':[14] X destroys Y, where X = Y. X = Y constitutes the center, while the excluded periphery encompasses the possible alternative "destroyers" or "destroyed". In other cases, the coreference occurs without a strong contrast (*Selbst·bewusstsein* 'self-awareness'), or the contrast occurs without the coreference component being evident or central (*Selbst·läufer* 'sure-fire success';[15] *auto·estinguente* 'self-extinguishing', *auto·sufficiente* 'self-sufficient').

Considering the prototypical part of this function, I therefore suggest the term **contrastive coreference** to refer to the function(s) of "reflexive word-formation markers". Consequently, I will call the latter **contrastive coreference markers** (henceforth $_{CC}$markers), and **contrastive coreference formations** (henceforth $_{CC}$formations)[16] will refer to the complex words built combining $_{CC}$markers with a variety of (mostly) deverbal base words.

[14] In the previous examples, it is worth noting how the use of a middle dot (<·>) is used as an orthographic means to explicit word-internal boundaries, especially to highlight the boundary between a contrastive coreference marker and its base. This is done to avoid the hyphen that could be confused with orthographic choices peculiar to a language. Compare, for instance, Italian *autodistruzione*, or German *Selbstzerstörung*, where the use of hyphen would be orthographically unusual, and English *self-destruction*, where the absence of the hyphen would be atypical.

[15] A *Selbstläufer* – lit. *self-runner* – is not a runner, but something that achieves the desired goal without much needed to be done to obtain it: «etwas, was wie von selbst, ohne dass viel dafür getan werden müsste, den gewünschten Erfolg hat» duden.

[16] I chose the rather unusual term "formation" here to avoid both "derivative" and "compound" and keep a neutral position in terms of category membership of the complex words with $_{CC}$marker to subdomains of word-formation.

2

THE BOUNDARIES OF WORD-FORMATION AND CONSTRUCTION MORPHOLOGY

In the previous chapter, we have introduced the topic of this study, and surveyed the literature dealing with the marking of reflexivity and the use of intensifiers in complex words. In this chapter, we will discuss the problems of the boundaries of word-formation, and introduce the theoretical framework which will permeate our discussion of the cross-linguistic data, i.e. Construction Morphology.

2.1 Form and function of contrastive coreference markers

From the survey of previous studies of what I have defined as contrastive coreference markers and formations, a series of dimensions of variation can be derived. Some concern formal aspects of the formations, while others concern their meaning.

Concerning formal aspects, we must remark that in the literature there is a rather limited interest about the morphological aspects of these complex words. The markers considered are either ascribed to prefixation or to compounding, with little discussion. However, several issues of morphological interest emerge in the domain of $_{CC}$markers. For example, it is not unproblematic to attribute a closed class element, e.g. German *selbst*, to compounding, whose productive patterns usually involve major lexical classes such as nouns, verbs and adjectives. Additionally, the fact that very similar word-formation meanings can be conveyed in different languages using different formal devices, is rather relevant to both the cross-linguistic study of word-formation and the issue of the boundary between derivation and compounding. In this context, it is important to note that in the languages discussed in the literature concerning $_{CC}$formations, $_{CC}$markers occur on the left side of the relevant complex words, never on the right side.

As far as the right side of $_{CC}$formations is concerned, we have already mentioned that the literature discusses mostly complex words in which the right constituent is a deverbal nominal – especially action nominals – with a few mentions of deverbal adjectives and some studies focusing on verbs combined with $_{CC}$markers. However, various possible types of base words are at best only listed, and any further implications for morphological theory are completely overlooked. What also stands out in this context, for example, is the boundary between derivation and compounding. If a $_{CC}$marker in a specific language belongs to the class of prefixes, and if it implies an operation on the argument structure of the underlying verb, then we expect that the prefixation operates first on the verb, and only

afterwards is a deverbal nominal or adjective derived on the basis of the prefixed verb. Conversely, if $_{CC}$formations in a specific language belong to the class of compounds, then they should be interpreted as a kind of synthetic compounds, where the $_{CC}$marker acts as internal argument of the underlying verb.

Both scenarios – that ideally characterize $_{CC}$markers in different languages, but it cannot be excluded that they can be relevant for a single $_{CC}$marker in a specific language – lead to further theoretical complications. On one hand, the idea that a prefixal $_{CC}$marker must first combine with verbs, and then we can obtain individual $_{CC}$formations from the prefixed verbs, runs counter to the observation that (action) nominals are the most productive domain of $_{CC}$markers.

On the other hand, synthetic compounds are well-known as a problematic phenomenon for word-formation theory (see Gaeta 2010), since they imply either the simultaneous application of different operations – e.g. the unproductive operation of forming NV compounds and the derivation of a deverbal noun or gerund – or unattested steps – e.g. *truckdriver* from °*truckdrive* 'to drive a truck', or *trendsetter* from °*setter* as 'one who sets something' – which are frequently subject to back-formation. Additionally, in the case of $_{CC}$markers, they cannot be simply assimilated with the modifier of a synthetic compound, precisely because intensifiers – which, as we have seen, constitute a privileged domain of origin for compound $_{CC}$markers – do not necessarily belong to the class of nouns, which is usually the class to which the modifiers of synthetic compounds belong.

A formal aspect can be relevant as far as the meaning of $_{CC}$markers and their contribution in $_{CC}$formations is concerned. We have seen above that in some languages different $_{CC}$markers compete, and in some cases display certain functional specialization. It appears that the literature has so far been primarily interested in $_{CC}$formations in which the right constituent is a deverbal derivative or a verb, but it has also been shown how competing $_{CC}$markers, such as German *eigen-,* have a wider compatibility when it comes to including underived base words. König (2011:123) briefly hints at the possibility that additional functions exist in this domain– «A vague possessive relation (*appartenance*)» – but to the best of my knowledge, no further discussion has been undertaken in the literature.

A further way in which competition between strategies surfaces in the realm of $_{CC}$formations concerns the cases in which complex words in one language correspond to (more or less) lexicalized phrases in another. For this possibility, König (2011: 114) also provides some compelling examples from French, in which

he considers the translation equivalents of English *self-love* (*amour-propre*), *self-loathing* (*mépris de soi*) and German *Eigeninitiative* 'one's own initiative' (*initiative personnelle*).

Let us consider them in more detail. In *amour(-)propre* 'self-love', the adjective and attributive intensifier *propre* 'own' modifies the noun *amour*. This first example is quite problematic from the point of view of the morphology-syntax boundary, since it is almost impossible to decide whether it must be conceived of as a product of morphology – as a left-headed Adj-N compound – or of syntax – as a complex noun phrase where the noun is modified by the adjective. In *mépris de soi* 'self-loathing', the noun *mépris* is modified by the stressed reflexive pronoun *soi,* embedded in a prepositional phrase. While this second case is formally less problematic, and can be safely considered a product of syntax, it remains debatable whether it is lexicalized or not. From the semantic point of view, the meaning of the phrase is rather compositional, so it need not be stored; however, we can argue that it has nonetheless been stored in the lexicon, although it does not display an idiosyncratic meaning.[17] Finally, *initiative personnelle* 'one's own initiative' involves the adjective *personnel* that conveys some kind of self-related meaning. As in *amour(-) propre* it is not clear whether it should be attributed to word-formation or syntax. While König, without much discussion, attributes *amour-propre* to word-formation and *initiative personnelle* to syntax – probably because the customary syntactic use of *propre* as an attributive intensifier is prenominal and it is accompanied by a possessive, as in *son (propre) amour* 'his/her (own) love' – this conclusion cannot be so easily taken for granted. For example, if we take into consideration that *un amour propre* contains the adjective *propre* 'clean', we can see the same relative order of adjective and noun in both *amour(-)propre* and *initiative personnelle*, i.e.the unmarked relative order of adjective and noun in the noun phrase in French.

We have devoted some space to the discussion of these potential competing strategies of French *auto-* in order to define the scope of this study. Albeit interesting, the competition of $_{CC}$formations with phrases or other products of syntax with comparable meanings, is not going to be taken into account.[18] On the one hand, we

[17] On the concept of lexicalization in the literature and on how the term will be used in this study, see Chapter 3, section 3.1.3.

[18] But see Angster (2017) for an account of possible historical forerunners of $_{CC}$formations in older stages of Italian, characterized by *di sé* 'of oneself' and *da sé* 'by oneself', i.e. the stressed Italian reflexive pronoun embedded in a prepositional phrase (compare with French *de soi* above).

would like to focus on (productive) word-formation patterns. On the other hand, accounting for such syntactic strategies might take us too far from word-formation, since it would imply accounting for a number of syntactic constructions that involve reflexive markers, intensifiers of various kinds, but also lexical items (such as the French adjective *personnel* above) which share a (maybe even marginal) function of $_{CC}$markers only in specific readings or individual expressions.[19]

We can now summarize the relevant issues concerning $_{CC}$formations and $_{CC}$markers which will direct our survey of these word-formation patterns in European languages. Some of them have been already pointed out in König (2011), while others are more closely related to the implications for word-formation. I will subdivide them in four groups, the ones relevant to $_{CC}$formations as a whole, the ones relevant to the left constituent (i.e. the $_{CC}$marker), the ones relevant for the right constituent (i.e. the base word), and finally, the ones concerning the meaning contribution of $_{CC}$markers in $_{CC}$formations.

(18) $_{CC}$formations as a whole:
- presence/absence of (productive) $_{CC}$markers in a language

(19) $_{CC}$markers:
- possible lexical sources (interaction with intensifiers and reflexives)
- membership within word-formation
- competition of different $_{CC}$markers

(20) base words within $_{CC}$formations:
- typology of base words

(21) meaning of $_{CC}$markers:
- functional specialization of competing $_{CC}$markers in a language

The issues in (18) and (19) will be the topic of Chapter 4. With the ones in (20) and (21), we will deal in Chapter 5 instead. Given the cross-linguistic perspective of this study, we will not be able to discuss in detail the aspect of the meaning

[19] This is also true of some affixes, e.g. Italian *intro-* in *introspezione* 'introspection, the act of looking into oneself', allowing for a certain reflexive reading which is completely absent in *introduzione* 'introduction', interpretable as 'the act of guiding someone into the understanding of a topic'.

contribution of $_{CC}$markers in all formations attested in the individual languages: this would require in-depth knowledge of the data and a richer empirical base for each language of the sample. Our intention here is to focus on the theoretical problem of the internal boundaries of word-formation, especially the one between derivation and compounding, and on how the cross-linguistic study of $_{CC}$markers can shed light on this issue. In order to deal with this main topic, we will now introduce the theoretical framework that will back up the analysis of $_{CC}$formations in the languages of Europe.

2.2 The boundaries of word-formation

The linguistic domain which constitutes the focus of this study is word-formation, and the theoretical problem which we will deal with is that of its boundaries. The discussion about the boundaries of word-formation involves exploring the relationship of word-formation with other domains that are not part of morphology – especially syntax – or with the other main subfield of morphology – i.e. inflection. But it also involves the exploration of the internal boundaries of word-formation, the most important of which is the distinction between derivation and compounding.

A discussion of the subdomains of word-formation and their interaction with adjacent areas cannot be carried out without referring to the units of word-formation and to the operations applied to those units. The nature of units and operations is also linked to the choice of a model of the lexicon that should define:

i. which units are part of the lexicon,
ii. whether affixes should be interpreted as units or as processes,
iii. if affixes are units, whether operations are also part of the lexicon or of the grammar,
iv. whether among the elements listed in the lexicon (listemes) we should include only the products of morphology or also the product of syntax,
v. only idiosyncratic or idiosyncratic as well as regular formations.

This list is undoubtedly incomplete, but it constitutes a sufficiently detailed roadmap to lay down the theoretical framework that we will adopt in this study.

I will first briefly discuss the issue of one of the external boundaries of word-formation: the boundary between inflection and derivation.

2.2.1 The boundary between inflection and derivation

The relationship between syntax and morphology is certainly a hotly debated issue, with scholars arguing either for the independence of (at least a part of) morphological phenomena from syntax, or for their complete reduction to by-products of syntactic operations. An especially animated area of discussion concerns the nature of inflection, first and foremost contextual inflection, i.e. the set of morphological exponents whose occurrence is controlled by agreement. The debate around the *split morphology hypothesis* – put forward in Perlmutter (1988), developed in Anderson (1992), and argued against in Booij (1994, 1996) – is an example of the advance of syntax in the territory of morphological phenomena. In an attempt at safeguarding the domain of morphology, Booij (1996) shows how the split morphology hypothesis, if accepted, would lead to the split of inflection itself, with contextual inflection (for example, gender and number marking of adjectives) ending up as part of syntax, and inherent inflection (not controlled by agreement, for example, number marking and gender of nouns) remaining part of morphology.

We will not further delve into the debate concerning the boundary between syntax and inflection, because it is not particularly pertinent to our discussion of $_{CC}$markers, but we will again come across an external boundary of morphology in our considerations of the boundary between compounding and syntax.

The distinction between contextual and inherent inflection is nonetheless helpful in introducing the topic of the boundary between inflection and derivation, which distinguishes the former from the wider domain of word-formation. As ten Hacken (2014: 10-11) notes, while both «are concerned with morphologically related forms», the core opposition between inflection and derivation consists in the type of relations that are established between those forms. He takes two examples as prototypical of this core distinction: the ten case and number forms of the Polish noun *kobieta* 'woman', and the English pair *read ~ readable*.

The first example is a prototypical case of inflection. The set of ten word-forms of *kobieta* are arranged in fourteen slots, which together form a paradigm that is characteric of most Polish nouns. Paradigm and citation form constitute the lexeme *kobieta*.

On the other hand, the second example of relation is taken as a prototypical case

of derivation. The English word *read* is a verb, whereas *readable* is an adjective – while all forms of *kobieta* are instances of a noun. In the same way, *read* does not form a paradigm together with *readable*, but they are two different lexemes, the latter being the result of the formation of a new lexeme.

The example highlights three fundamental aspects of the distinction between inflection and derivation, i.e. the structuration of inflection in paradigms of word-forms, the homogeneity of paradigm forms from the point of view of their syntactic category, and the ability of derivation to establish relations between different lexemes.

As the term prototype used by ten Hacken (2014) suggests, the distinction between inflection and derivation is raveled by a wealth of borderline cases. Arkadiev and Klamer (2019: 442) note that the literature regards the two prototypes as poles of a continuum. A number of features – including two of the three mentioned above, i.e. "function" and "paradigmaticity" – have been identified as useful for modelling this continuum, and Table 1 below lists the ones Arkadiev and Klamer (2019: 443) picked out from the recent literature.

Table 1. Features of prototypical inflection and derivation – (Arkadiev and Klamer 2019: 443)

Parameter	Inflection	Derivation
Function	Does not change syntactic category of a word	May change syntactic category of a word
Meaning	Often has purely grammatical meaning	Tends to have lexical semantic content, i.e. meanings similar to the meanings of independent words
Regularity	Is often semantically regular	May have unpredictable semantic content
Syntactic determinism	Is often syntactically determined	Does not require a specific syntactic environment
Obligatoriness	Function is obligatory	Function is not obligatory
Productivity	Is highly productive	Often applies only to certain words, or classes of words
Paradigmaticity	Is often organized in paradigms	Is often not organized in paradigms
Fusion	Can be marked by portmanteau morphemes	Is rarely marked by portmanteau morphemes
Recursivity	Is marked only once in the same word	May apply twice in the same word

Under "syntactic determinism", we can recognize the features relevant in distinguishing contextual and inherent inflection. The table lacks a clear feature referring to the creation of new lexemes typical of prototypical derivation, but a couple of other functions point in the direction of aspects relevant for the lexicon. The features "meaning" and "regularity of meaning" contrast inflection, which conveys a grammatical, predictable meaning, with derivation, whose meaning contribution tends to be lexical, i.e. «similar to the meanings of independent words», and idiosyncratic, which means that it might be stored independently of the regularities obtained from comparing the meaning of series of related forms, a fact that resembles again the behavior of simple, non-derived lexemes, rather than that of individual forms in a paradigm.

A similar semantic contrast is exploited by Bauer (2004), in the context of an alternative proposal, to deal with the boundary between inflection and derivation. Instead of using features to define prototypes, he identifies six classes of morphological processes (Table 2), challenging the traditional idea that only two recurring clusters of features – inflection and derivation – are to be recognized.

Table 2. Six-way classification of morphological processes between inflection and derivation – (Bauer 2004)

<table>
<tr><th>Contextual</th><th>Inherent</th><th>Valency-Changing</th><th>Transpositional</th><th>Evaluative</th><th>Lexicon-expanding</th></tr>
<tr><td>Agreement</td><td colspan="5">No agreement</td></tr>
<tr><td colspan="2">Lexeme-maintaining</td><td colspan="4">Creating new lexemes</td></tr>
<tr><td colspan="3">Class-maintaining</td><td colspan="3">Class-changing</td></tr>
<tr><td colspan="4">Grammatical</td><td colspan="2">Lexical</td></tr>
<tr><td colspan="5">Paradigmatic</td><td>Non-paradigmatic</td></tr>
</table>

The classification is not unproblematic, and Bauer himself is well aware of its issues.[20]

The features in Table 1 and the classes in Table 2, suggested by Bauer (2004), are interesting in the context of our discussion of $_{CC}$markers, because they both

[20] «Evaluative morphology is awkwardly placed here, since it is typically class-maintaining, though it can be class-changing.» (Bauer 2004: 283), and, further below: «The creation of a new set of categories where there have previously been only two or three inevitably brings problems with it. There are two obvious problems with the suggestion made here: (1) it creates more borderlines, and thus more borderline disputes; (2) it splits the outputs of more morphological processes into different categories.» (Bauer 2004: 284)

show how prototypical inflection/contextual inflection is characterized by grammatical meanings, while prototypical derivation and Bauer's classes of evaluative and lexicon-expanding morphology are characterized by lexical meanings. As we have seen above, the contribution of $_{CC}$markers to $_{CC}$formations consists not in lexical, but in grammatical meaning, which has been compared with the meaning of reflexive pronouns, intensifiers or of the anticausative alternations. Reflexivity and anticausativity can be considered as valency-changing functions, which in Bauer's model co-occur with grammatical meaning and the creation of new lexemes, so that $_{CC}$markers can be interpreted as an instance of some non-prototypical form of derivation in the feature-based model in Table 1 above.

However, two issues emerge. The first concerns the contrastive value of $_{CC}$markers, a pragmatic component of their interpretation, which on the one hand does not fit into the valency-changing class, while on the other hand, it represents an example of "fusion", i.e. of cumulative exponence,[21] an aspect which is usually typical of inflection.[22] The second, more relevant issue concerns the fact that some $_{CC}$markers have been regarded as compounds, since they involve two elements that can occur as independent, non-bound words. On the one hand, this shows that also compounds can apparently convey a grammatical meaning instead of lexical. On the other hand, this suggests that the continuum between inflection and derivation should be regarded as a continuum between inflection and word-formation, therefore not as a linear continuum, but at least as a planar continuum that also involves compounding.

2.2.2 The boundary between derivation and compounding

This leads us to the boundary between derivation and compounding. If we consider again Table 1, derivation and compounding share more or less the same features which contrast derivation with inflection, while the prototypical form of compounding can be regarded as part of the lexicon-expanding class proposed by

[21] See Ricca (2005) on examples of cumulative exponence in derivation.

[22] One could argue that inflected forms can also be characterized by context-related, sociolinguistic or even pragmatic restrictions of occurrence – see, for example, Thornton (2011) on conditions constraining the occurrence of morphological doublets ("cell-mates" in Thornton's terms). However, the constraints considered by Thornton appear to apply to whole inflected forms and not just to the relevant affixes or morphological operations.

Bauer (2004). Some additional discussion of the features is nonetheless necessary.

As we have already mentioned, the meaning of derivational processes can also be very grammatical, especially in the case of transpositional and valency-changing morphology. As noted above, $_{CC}$markers might constitute a similar case in compounding.

If we consider regularity, meaning regularity is higher in low-frequency products (or even nonce-formations) of highly productive morphological operations. Conversely, it is lower in the case of high-frequency products of morphological operations, and it increases as the productivity of an operation decreases.

The regularity of meaning is thus closely connected with productivity, i.e. with the characteristic of different morphological operations to display differences in availability and profitability, as defined by Bauer (2001). Derivational operations can be very productive, to the point of bordering on inflection in cases of transpositional morphology (see the status of Italian *-bile* bordering with inflection; Ricca 2004). However, it has been suggested that while derivational processes have a gradual productivity from very to weakly productive, compounding patterns, if they are productive at all, have an almost unrestricted productivity, similar to that of syntactic constructions (see Koefoed and Van Marle 2001: 304).[23] At any rate, both compounding and derivational structures can be completely unproductive (see the case of unavailable operations like English deadjectival nouns in *-th* and cranberry morphs). On the other hand, very productive semi-instantiated compositional operations are not a rare occurrence, but their high productivity is considered a criterion for membership in the intermediate class of affixoids (see below 2.2.6).

Also, a certain connection can be established between the features of paradigmaticity and fusion. On the one hand, in recent years there have been attempts at applying a paradigmatic model beyond the limits of inflection, aimed chiefly at including derivational operations, but also some compound structures (see Hatout and Namer 2019 and Gaeta and Angster 2019). On the other hand, it appears that recognizing cumulative exponence in derivation is more a matter of general modelling in this realm – i.e. the lack of a clear paradigmatic structure of derivation, such as exemplified by inflectional paradigms – than a problem of the lack of fu-

[23] «[T]he productivity of compounding resembles that of syntactic constructions: if a given compound type is productive, its productivity is of the kind we meet in syntax, which means that there are virtually no restrictions.» (Koefoed and Van Marle 2001: 304)

sion of the different meanings conveyed by a morphological operation. Given that many derivational operations can be recognized as competing with those that are more compound-like, it can be assumed that paradigmaticity and fusion cannot be excluded from compounds, if a model for the semantic content of word-formation operations is proposed.

Finally, while recursivity characterizes both derivation and compounding, as opposed to inflection, it is more constrained in derivation than in compounding, and more constrained in compounding than in syntax, while constraints vary greatly in different languages (Gaeta 2015: 864).[24]

So, if for most of the features considered above the difference between derivation and compounding is more a matter of degree and not of presence/absence of individual features, what distinguishes derivation from compounding? Considering again prototypical cases of the two classes, the answer to this question lies in the units involved.

Olsen (2014) first provides a definition of the two classes before turning to problematic phenomena.

> «Compounding, simply spoken, is a combinatorial word-formation process that creates complex words by combining lexemes (roots or stems). Its products, that is, compounds, are comprised of two or more lexemes at the word level such as *cheek bone*. Compounding is most often contrasted with overt affixation which derives a word from a lexeme by adding an affix, that is a bound morpheme that combines with a specific category base to form a pattern. An example of suffix derivation with a simple lexeme as a base is *wire+less.*» Olsen (2014: 26)

Olsen also highlights some commonalities shared by the two classes, but with the only effect of strengthening the relevance of the units involved in establishing the boundary between them:

[24] We can suppose that more restrictive constraints on combinations apply in derivation than in compounding. See Hay and Plag (2004) on constraints on suffix combinations in English, and Manova (2015) on closing suffixes. Concerning compounds, positionally bound morphemes (Olsen 2015: 375) – such as the elements from the disputed class of affixoids bordering on derivation; see 2.2.6 below – appear as rather constrained. Further research and a clearer modelling of restrictions in compounding is nonetheless needed.

> «[A]ffixation and compounding share most of their formal properties: they are binary branching, recursive, headed structures. Especially in languages that have right-headed compounds, like Germanic languages, the primary difference between affixation and compounding lies in the status of the constituent parts: if at the relevant level of analysis both constituents are lexemes belonging to the open word classes of the language, the result is a compound, if one constituent is a formative, that is a bound morpheme belonging to a finite class of elements in the language, the structure is an affixation.» Olsen (2014: 27)

A number of concepts are used in the definitions above, whose definitions in turn are far from unproblematic, starting from that of "lexeme" and its status, compared to that of "word", "root" and "stem", as opposed to the concept of "bound-morpheme". It is also worth noting how Olsen resorts to the distinction between "open word classes" and "finite classes of elements" for contrasting compound and derivational units. Without providing a survey of the definitions of these concepts found in the literature, I will now briefly discuss them.

2.2.3 The word level and below

The first concept to be defined is that of word, and the distinction between word and lexeme.

> «One well-known definition of the word runs as follows: 'A word may be defined as the union of a particular meaning with a particular complex of sounds capable of a particular grammatical employment.' This definition, it will be observed, makes it a necessary condition that the word should be simultaneously a semantic, a phonological and a grammatical unit.» (Lyons 1968: 200)

Lyons' definition of word as a multilayered unit combining semantic, phonological and grammatical features appears very similar to the following definition of lexeme, which Montermini (2010) provides in the context of a discussion of the units in compounding:

> «In most word-based theories of morphology the lexeme defines as an elementary sign combining at least three sets of properties, corresponding to three types of linguistic information: a phonological representation, a semantic representation, and morphosyntactic information (e.g. grammatical category).» (Montermini 2010: 86-87)

Another definition can help narrow down the scope of the concept of lexeme, as opposed to that of word.

> «In lexical analysis, lexeme identifies the unit of lexicon that constitutes the base form to which all the inflected forms are ascribed (for example, the conjugated forms of verbs or the inflected forms of nouns) and **lemma** corresponds to the lexeme seen as an entry in a dictionary.» (Ježek 2016: 22-23; emphasis in the original text)

Ježek's definition, closer to lexicology, points in the direction of the lexeme as a more abstract entity, which constitutes the common thread linking more concrete embodiments, e.g. the inflected forms of elements of classes like nouns, verbs, or adjectives in inflectional languages.

We will consider lexemes as more abstract units stored in the lexicon along with families of (grammatical) words, i.e. paradigms. The concept of "grammatical word" can be defined as follows:

> «A *grammatical* word consists of a number of grammatical elements which (i) always occur together, rather than scattered through the clause (the criterion of cohesiveness); (ii) occur in fixed order; and (iii) have a conventionalized coherence and meaning (Dixon and Aikhenvald (2002).» (Aikhenvald 2007: 2; emphasis in the original text)

Following Aikhenvald's definition, grammatical words – which constitutes only one of the possible definitions of word that are surveyed and discussed in Dixon and Aikhenvald (2002) –[25] appear as more concrete entities compared to

[25] Among the different possible types of word, a prominent position belongs to the phonological word (Dixon and Aikhenvald 2002: 13-18), whose definition is often instrumental in the distinction between words and clitics (Dixon and Aikhenvald 2002: 25), while stress patterns are used to distinguish

lexemes. When we mention words in this study, we will keep in mind this lower level of abstraction. This does not mean that lexeme and word cannot overlap in some context, especially in cases in which a language does not have a rich inflection, but even in those cases, it is always possible to set apart the lexeme as a unit of the lexicon from the word as the embodiment of a lexeme, especially as a unit holding a specific syntactic (or morphological) function, irrespective of whether this implies a modification in its form or not.

Before continuing our survey of the relevant terminology considering units below the level of the (grammatical) word, it is worth mentioning two aspects of the perspective we will adopt in this study. On the one hand, it is important to stress that we will maintain a word-based approach to word-formation, and, accordingly, consider affixes and units below the word level as non-independent from morphological operations – similarly to how they are regarded in word-based models of lexicalism (e.g. Aronoff 1976).[26] We might, however, resort to a morpheme-based terminology for descriptive purposes in the discussion of the cross-linguistic data.

On the other hand, while we do not support the morpheme-based idea that affixes and, more generally, morphemes are independently stored in the lexicon, from our perspective, the lexeme is not the only type of entity stored in the lexicon: we support a view of the lexicon in which elements bigger than the (grammatical) word, i.e. products of syntax, can also be stored. Following Di Sciullo and Williams (1987), we will call the set of objects memorized in the lexicon listemes, be they lexemes/words or phrases/multi-word units (see the discussion on the boundary of compounding and syntax below in 2.2.8).

Following Bauer (2014),[27] below the word level we can recognize roots, stems, bases – all terms which apply to the part of a word holding the fundamental lexical meaning – and affixes. To summarize, while the root is the «smallest core of any word», the stem is a form of the lexeme that cannot occur as a word-form (but which attaches to inflectional affixes, if we follow Bauer 1983: 20). The term

between cohering and non-cohering derivational affixes (Raffelsiefen 2015: 900). Concerning compounds, phonological criteria are very sensitive to language-specific phonological features (see Lieber and Štekauer 2009a: 11).

[26] For a survey of Lexicalist Morphology and its different theoretical nuances see Montermini (2019).

[27] Bauer (1983: 20-22) introduces the terms root, stem and base in a way very similar to Bauer (2014), but he restricts the use of stem to the inflectional domain. Although Bauer (2014) adopts the wider sense of stem, we will stick to the more restricted sense found in Bauer (1983).

base crosses the distinction between root and stem in the sense that both roots and stems can constitute the base of a complex word, i.e. the form to which an affix – defined as «a recurrent piece of phonological material, not itself a root, which when found in a word has a relatively consistent effect on the meaning of the word in which it is found» – attaches (Bauer 2014: 118). Another relevant distinction in this realm is that between bound and free morphs. Affixes are bound morphs, they «cannot stand alone as utterances» (Bauer 2014: 119), however, if the stem cannot occur as a word-form, the stem can also be bound, and a stem which is also a root is a bound morph as well.

We have hinted at the fact that roots, stems and bases are all linked to a lexeme, while affixes are not since they are not roots. The circularity of this statement is evident, and compels us to look deeper into the distinction between roots and affixes. It is clear that some languages display mostly free roots – i.e. they can occur as word-forms in a sentence, see (22).a – while in other languages roots are usually bound – they constitute one of the stems to which inflectional affixes attach, see (22).b.

(22) a. English *cat* 'cat.SG' ~ *cat-s* 'cat-PL'
b. Italian *gatt-o* 'cat-SG' ~ *gatt-i* 'cat-PL'

The word-form for 'cat.SG' formally corresponds to the root *cat*, while in Italian the root *gatt-* can never occur as a word-form. Beyond the different ways in which the two languages mark number distinctions, in this example it is clear how the lexical content in both languages is conveyed by the root and not by the suffixes, whose contribution is more grammatical, bearing the values of the category number. This is not unexpected since the example shows a case of inflectional morphology.

Things become a bit more complex if we move from inflection to derivation, because according to Arkadiev and Klamer's (2019) features, derivation should convey «lexical meaning [ ] similar to that of independent words», while in Bauer's (2004) classification of morphological classes, although valency-changing and transpositional morphologies have grammatical meaning, evaluative and lexicon-expanding morphologies are considered to have lexical meaning. Therefore, we have both roots and derivational affixes which can have lexical meaning and occur as bound morphs. On the other hand, we have free morphs, such as function

words – prepositions, auxiliaries, pronouns, etc., i.e. closed-class elements – that have grammatical meaning, but occur as independent grammatical words.

This makes for a rather intricate picture, which can probably be clarified by shifting attention from semantic to formal aspects. Affixes, unlike roots, regardless of the type of contribution they convey, rely on a base, which means that the base is essential for an affix to occur.[28] Conversely, roots can occur without the presence of affixes – at least of derivational ones, in languages with (rich) inflection. We can thus distinguish between (prototypical) roots and affixes based on this formal aspect.

2.2.4 Borderline units: neoclassical combining forms

However, if we go back to the boundary between derivation and compounding, we will find several classes of phenomena which involve units that are half-way between the two. The most relevant to our discussion of $_{CC}$markers are neoclassical combining forms and affixoids (also called semi-affixes).

Iacobini (2004a) offers a survey of neoclassical combining forms in Italian, enumerating a series of their features that holds true for other languages in which they are used (see, e.g. Olsen 2019: 35-36). Neoclassical combining forms do not occur autonomously in the sentence, and are called neoclassical since they are mostly taken from classical languages (mostly Ancient Greek and Latin). Being widespread with similar form and meaning in the lexicon of many languages – especially in technical-scientific terminology of European languages – they are also called "internationalisms". These characteristics of neoclassical combining forms make them a rather homogeneous class from a cross-linguistic formal point of view as well: despite the fact that they are always bound, each of these elements can occur either as the first (generally with modifier function), or as the second element (generally with head function) in relevant formations, behaving similarly to common compound elements. They convey lexical meaning, and are often used in complex formations as an alternative to free words of the same meaning

[28] From a word-based perspective, this is the essence of word-formation rules, introduced in the lexicalist paradigm by Aronoff (1976): «a [word-formation rule] can be viewed as a means of establishing a correspondence between the phonological, semantic and syntactic (primarily categorical) properties of a simple(r) and of a complex word» (Montermini 2019: 126). Aronoff's (1976) perspective concerning units in word-formation is word-based, with the consequence that an affix «cannot be separated from the rule, because it is nowhere given any representation on its own» (Aronoff 1976: 70).

(*-lito-* 'stone', *-zoo-* 'animal'). In addition, like words and unlike affixes, they can be derived (*cerebro-* 'brain' → *cerebrale* 'concerning the brain; cerebral', *etno-* 'nation, people' → *etnico* 'belonging to a population, ethnic', *ipno-* 'sleep' → *ipnosi* 'hypnosis'). Finally, resembling words in this respect as well, neoclassical combining forms constitute an open class, while affixes in synchrony tend to constitute a closed class of elements (Iacobini 2004a: 69, 84-85).

Guevara and Scalise (2008: 111) have highlighted how neoclassical combining forms can also occur in compounds with free words (23), and that in these formations the relationships between constituents do not differ from the relationships that can occur between the constituents of compounds formed exclusively of words.[29]

(23)	English:	anthropology	[sW+sW]
		biochemistry	[sW+W]
	Italian:	odontotecnico	[sW+W] = 'lit. tooth technician'
		colorificio	[W+sW] = 'lit. paint factory'

For this reason, they conclude that neoclassical combining forms should not be treated differently from "normal" words (Guevara and Scalise 2008: 111).

Iacobini (2004a) also remarks how another terminological tradition has emphasized the affinity of neoclassical formative elements with affixes, dividing them into "prefixoids" and "suffixoids" (see *affissoidi*/*prefissoidi*/*suffissoidi* in Migliorini 1963).[30] In fact, not all neoclassical combining forms can occur indifferently as the first or second element: some of them are used exclusively in initial or final position. The elements present exclusively in the final position (e.g. Italian *-cida*, *-crate*, *-fero*, *-geno*, *-voro*, etc.) generally form nouns and adjectives of agentive and instrumental meaning in formations that parallel synthetic compounds (*carni-voro* ≈ *meat-eater*), or Romance VN compounds (*vessilli-fero* 'flag-bearer' ≈ *porta-bandiera* 'flag-bearer, lit. bear-flag' Iacobini 2004a: 89).

The elements that always occur in the first position, on the other hand, have characteristics similar to prefixes, and express a relational meaning that indicates

[29] Scalise (1994: 269-271) uses the term *semiparola* 'semi-word' to refer to neoclassical combining forms. In example (23), W stands for "word" and sW for "semi-word".

[30] This terminology is somewhat problematic since the same terms have been used to identify a rather different class of phenomena, see further below in 2.2.6.

position (*ecto-*, *endo-*, *exo-*, *meso-*, *peri-*), temporality (*archeo-*, *vetero-*), quantity (*hemi-*, *equi-*, *omni-*, *panto-*) and number (*uni-*, *mono-*, *bi-*, *tri-*). Finally, some elements of this group (*auto-*, *iper-*, *ipo-*, *macro-*, *mega-*, *meta-*, *micro-*, *multi-*, *neo-*, *para-*, *pluri-*, *poli-*) are considered by Iacobini to be part of the class of prefixes, based on the observation that they are used «before words in numerous formations of common usage, and the type of meanings they express triggers processes of grammaticalization» (Iacobini 2004a: 88; my translation).[31]

The fact that neoclassical combining forms are problematic units from the point of view of their formal correspondence with free words challenges their classification as words, lexemes or affixes, and the definition of compounds, as far as the elements allowed and the processes underlying their formation are concerned.

However, as highlighted by Montermini (2010: 82), cases of formally problematic units are not so problematic if the lexeme is assumed to be an abstract unit of lexical organization (24.b), and if problematic units are conceived of as instances of divergence between the form taken by a lexeme in a compound (24.c), and the word forms available for that lexeme (24.d), which is why they can be treated as particular cases of word-based composition.

(24) a. Fr. *ludothèque* 'toy library', It. *cardiologo* 'cardiologist'

b.	lexeme:	Fr. JEU	It. CUORE
c.	[+ bound]:	*ludo-*	*cardio-*
d.	[– bound]:	*jeu*	*cuore* (adapted from: Montermini 2010: 82)

The case of neoclassical combining forms – which, as Montermini remarks, are assimilable to analogous phenomena present outside the context of Western Europe or of elements of neoclassical origin – can be considered as lexical elements characterized by the feature [+ bound], and according to Montermini (2010: 82), this is not in contrast with the idea that lexemes (and not morphemes) are the units according to which the mental lexicon is organized.

[31] «L'impiego davanti a parole in numerose formazioni di uso comune e il tipo di significati espressi hanno provocato processi di grammaticalizzazione che a nostro avviso giustificano l'inserimento nel novero dei prefissi di elementi come *auto-* "da sé", *iper-*, *ipo-*, *macro-*, *mega-*, *meta-*, *micro-*, *multi-*, *neo-*, *para-*, *pluri-*, *poli-*.» (Iacobini 2004a: 88)

2.2.5 Borderline units: prefixes at the boundary with compounding

It is worth noting that Iacobini (2004a, 2004b) – but also Montermini (2008) – include Italian *auto-* among Italian prefixes, despite its neoclassical origin and formal properties. We have also observed that the other $_{CC}$markers considered in the literature occur on the left of the relevant $_{CC}$formations, to the point that some of them are also regarded as prefixes (e.g. English *self-* in Mackenzie 2018). Before we turn our attention to affixoids, it would be helpful to briefly outline the boundary between prefixation and compounding.

Iacobini (2004b: 105-107) defines the features that are typical of prefixes by contrasting them to compounds (and neoclassical combining forms). According to him, the most important characteristics that allow prefixes to be seen as derivational affixes and to distinguish them from words are:

i. the fixed position within a complex word,
ii. the impossibility of being derived or inflected,
iii. the relationship (exclusively determinative [i.e. never coordinative]) with the base.

Other criteria proposed as relevant are:

iv. the relational semantics expressed by the prefixes,
v. the position of the head of the prefixed word (on the right, like in suffixed words, instead of on the left, like in many productive compounds [in Italian]),
vi. the fact of constituting a tendentially closed inventory,
vii. the fact of being bound elements,
viii. the fact of being elements on average shorter than words.[32]

However, it is well known that prefixes – at least in European languages like Italian, French, but also German – frequently correspond with prepositions and adverbs. Until recently, this has lead a number of scholars in the tradition of de-

[32] «Le caratteristiche più importanti al fine di considerare i prefissi come affissi derivazionali e di distinguerli dalle parole sono: la posizione fissa all'interno della parola complessa, l'impossibilità di essere derivati o flessi, la relazione (esclusivamente di tipo determinativo) con la base; altri criteri rilevanti sono la semantica di tipo relazionale espressa dai prefissi, la posizione della testa della parola prefissata (a destra, così come le parole suffissate, anziché a sinistra come molti dei composti produttivi), il fatto di costituire un inventario tendenzialmente chiuso, di essere elementi legati, di essere elementi mediamente più brevi delle parole.» (Iacobini 2004b: 107)

scriptive grammar to consider prefixes as an instance of compounding.[33] This account is nonetheless to be dismissed. Montermini (2008: 32) points out that because of their higher phonological and syntactic autonomy, (Italian) prefixes bear more similarities than suffixes to the lexical units from which they are derived through grammaticalization, but it is nevertheless possible to argue for the inclusion of prefixes into the class of affixes based on several observations, including the fact that the semantic contribution of prefixes and their homonymous prepositions do not overlap, and the range of meanings conveyed by prefixes is usually larger than that of prepositions (Montermini 2008: 34). Furthermore, prefixes differ from prepositions from a formal point of view: while prepositions take nouns as complements, prefixes can be combined with any major lexical category (nouns, verbs, adjectives; Montermini 2008: 35; Iacobini 2004b: 103).[34]

Interestingly, although the Italian *auto-* is considered a prefix by both Montermini (2008: 61, as "modal" prefix along with *co(n)-* and *ri-*)[35] and Iacobini (2004b: 159, «the only prefix expressing a reflexive value»), based on its high productivity, its fixed position on the left of the word and its compatibility with native words (*auto·valutazione*), it still shares aspects of a neoclassical combining form. For example, *auto-* still occurs with combining forms (*auto·psia*, *auto·bio·grafia*), can be derived (*aut·ismo* 'autism', *aut·ista* 'autistic person'; see *idro-* → *idr·ico* 'hydric'), and has some compatibility with coordinative structures, where it can occur coordinated with another modifier in a larger subordinative structure ([[*auto·etero*]·*aggressività*] 'self- and other-directed aggressiveness'), but not in a binary structure (compare e.g. *(equilibrio) idro·salino* '(balance) of water **and** salt'). We suppose that *auto-* is nonetheless included in the inventory of Italian prefixes, because it shares a behavior similar to that of other neoclassical combining forms such as *iper-*, *macro-*, *micro-*, *multi-*, *neo-*, *para-*, *pluri-*, which are considered prefixes because «they are part of the competence of most speakers, they combine with [native Italian] words

[33] See Bauer (2005: 106), Montermini (2008: 13-14), Olsen (2019: 36) for further references of this tendency in the French, Italian and German grammatical traditions.

[34] It must be added that some classes of exocentric compounds with a preposition acting as modifier (P+N: nouns: *senza·tetto* 'homeless (person)', *sotto·scala* 'space under a staircase', *dopo·lavoro* 'workers recreational club'; adjectives: *dopo·sci* 'après-ski', *raso·terra* 'close to the ground'; adverbs: *sotto·braccio* 'arms linked', *contro·voglia* 'reluctantly') are recognized by some scholars (Scalise 1983: 146-147; examples from Bisetto 2004: 47, 48, 51-52).

[35] Montermini (2008) provides a semantically driven classification of Italian prefixes that is applied by Lefer and Cartoni (2011) in a contrastive study of prefixes. An example of a more fine-grained classification of word-formation patterns is an areal-typological survey of valutative morphology in Grandi (2002).

in a rich series of neologisms in current use, in formations structured in the same way as those in which other prefixes of older attestation occur, with which they display paradigmatic relationships» (Iacobini 2004b: 101; my translation).[36]

If we turn our attention to German and to the boundary between derivation (especially prefixation) and compounding, the situation is similar to that of Italian, in that both languages display productive prefixation and also borderline phenomena. However, German differs from Italian in the number of prefixes and the lexical category of the bases to which they attach, but also in the type of borderline phenomena occurring in this language.

German affixes can be generally subdivided into two groups: native (*indigene Affixe*, Fleischer and Barz 2012; *native Affixe*, Eisenberg 2013) and non-native affixes (*Fremdpräfixe*, *Fremdsuffixe*, Fleischer and Barz 2012; *Lehnpräfixe*, *Lehnsuffixe* Eisenberg 2013).[37] Focusing on prefixes, it must be noted that native prefixes which combine with nouns (*erz-*, *ge-*, *haupt-*, *miss-*, *un-*, *ur-*) and adjectives (*erz-*, *un-*, *ur-*) are less numerous than:

i. the non-native prefixes which combine with nouns and adjectives (*anti-*, *contra-*, *dis-*, *ex-*, *hyper-*, *in-*, *inter-*, *meta-*, *neo-*, *non-*, *post-*, *prä-*, *super-*, *supra-*, *trans-*);
ii. the native prefixes which combine with verbs
    a. be-, *ent-*, *er-*, *ge-*, *miss-*, *ver-*, *zer-*;
    b. durch-, *hinter-*, *über-*, *um-*, *unter-*, *wider-*;
iii. the native suffixes taking nouns and adjectives as bases.[38]

With regards to the list of prefixes provided above, two facts are worth noting. On the one hand, among German non-native prefixes Eisenberg includes some

[36] «Riteniamo che elementi come *iper-*, *macro-*, *micro-*, *multi-*, *neo-*, *para-*, *pluri-*, possano essere a tutti gli effetti considerati prefissi in ragione del fatto che, sebbene siano impiegati anche in combinazione con elementi formativi nella coniazione di termini tecnico-scientifici, sono ormai entrati a far parte della competenza della generalità dei parlanti, e sono premessi a parole in numerosi neologismi d'uso corrente in costruzioni identiche a quelle a cui parteci- pano i prefissi di più antica attestazione in italiano, e con i quali intrattengono rapporti di tipo paradigmatico.» (Iacobini 2004b: 101)

[37] Eisenberg (2013: 36) differentiates between *Lehnwörter* 'loanwords', which do have a non-native origin, but are nonetheless integrated into German, and *Fremdwörter* 'foreign words', which have not fully adapted to the German language structure.

[38] Enumerating the suffixes would make the list too long: see Fleischer and Barz (2012: 56-57) for the complete list of native affixes, and Eisenberg (2013: 236 ff.) for the distinction between native and non-native prefixes.

elements that he also mentions when discussing neoclassical combining forms (*inter-*, *super-*, *trans-*). On the other hand, the list of verb prefixes under ii. is divided into two parts: the first one, under ii.a, includes the verb prefixes that have no free counterpart (preposition or adverb); the second one, under ii.b, includes the verb prefixes that have prepositions as free counterparts, but also compete with homonymic separable preverbs (see below).

These two aspects point in the direction of two borderline phenomena in the domain of German word-formation: the so-called *Konfixkomposita* and the *trennbare Verben*, also called separable complex verbs (SCVs).

German neoclassical combining forms (or *Konfixkomposita*) share many features that we have already discussed concerning Italian neoclassical combining forms – bound forms, absence of a free counterpart, lexical meaning, fixed position for subsets of combining forms (i.e. *Präkonfixe* and *Postkonfixe*), possibility for some of them to combine with native words, open class (see Eisenberg 2013: 231-235 for more details) – so we are not going to delve further into these considerations. It might nonetheless be relevant to mention only that German *auto-,* unlike Italian *auto-*, does not combine with native words (Fleischer and Barz 2012: 172), rather, it is limited to combining with other so-called *Konfixe*. It must be also added that in German, as well as in Italian, a free grammatical word exists – Ger. *Auto* from Ger. *Automobil*, It. *auto* from It. *automobile* – which, with the meaning 'car', can combine in both languages productively with native bases: Ger. *Auto·bahn* 'motorway', *Auto·besitzer* 'car owner'; It. *auto·strada* 'motorway', *auto·concessionario* 'car dealership'. In Italian, the bound use of *auto(-)* 'car' can sometimes overlap with that of the $_{CC}$marker: *auto·demolizione* usually means 'car scrappage', but it can be attested in the sense of 'self-destruction, demolition of one's own beliefs'.[39] This kind of elements, deriving from the clipping of neoclassical formations, homonymic with neoclassical combining forms, and often displaying a free counterpart and a higher degree of compatibility with native words, have been defined as "second generation" neoclassical combining forms (*elementi formativi neoclassici di seconda generazione*, Iacobini 2004a: 73-74).[40]

[39] See, for example, *Zidane continuava l'autodemolizione del proprio mito* 'Zidane carried on with the demolition of his own myth'. Retrieved from the page https://digilander.libero.it/mondojuventus/9900/news99.html; 16th February 2022.

[40] For a discussion of the "second generation" neoclassical combining forms and for more examples of Italian *auto-* 'car' in this use see Iacobini (2004a: 74-75).

2.2.6 Borderline units: affixoids in the German morphological tradition

Before turning our attention to separable complex verbs in 2.2.7 below, it is worth surveying another borderline case between derivation and compounding which is particularly relevant in the German morphological literature: affixoids.

We have seen above in 2.2.4 that in the Italian literature one of the terms used to label neoclassical combining forms is "affixoid". This term stresses the fact that these elements have features in common with affixes (fixed position, lack of formal correspondence with a syntactically free element), but they also share features with words, in particular because they carry a lexical meaning. The advantage of this term is its neutrality with respect to the etymological origin in the lexicon of classical languages.

The same term is used in the German tradition of morphological studies to indicate intermediate elements between compounding and derivation which have native origin. However, the crucial difference between neoclassical combining forms and German affixoids is that the features shared with affixes on the one hand, and with compound units on the other, are to some extent reversed. Like affixes, German affixoids are characterized by a rather abstract meaning, but their form actually corresponds to that of independent words.

The concept of affixoid has often been accompanied by a huge discussion between proponents/supporters and critics. For instance, ten Hacken (2001: 355-356) considers the case of affixoids from the point of view of the validity of this concept as an intermediate class, independent of composition and derivation. Ten Hacken notes how this proposal has been especially successful in the German research tradition between the 1970s and 1980s. The proposal originates from the observation of the existence of elements that have a transitional status between roots and affixes. See, for example, the name *Zeug* 'thing', in *Flugzeug* 'airplane', litt. 'flying thing', or *Schreibzeug* 'writing instruments'. The phenomenon is not limited to nouns, see, for instance, the adjective *frei* 'free' in *fehlerfrei* 'perfect', litt. 'error free', or *widerspruchsfrei* 'without opposition' (see Fleischer 1969: 63-66).

In a critique of the concept of affixoid, concerning its necessity and theoretical legitimacy, Schmidt (1987) indicates various criteria gradually proposed to isolate the class of affixoids; ten Hacken (2001) mentions three of them: (i) increased productivity, (ii) less semantic specificity, (iii) a formal or etymological link with

a free lexeme (Schmidt 1987, quoted in ten Hacken 2001). If the last criterion distinguishes affixoids from affixes, the first two are elements which allow affixoids to be distinguished from normal compounding and from the lexemes to which they correspond.

Ten Hacken notes how the case of affixoids has been used in item-and-arrangement morphological approaches to highlight the adjacency of derivation and composition, while item-and-processes approaches divide individual cases between the two domains of word-formation. Furthermore, he observes that after Schmidt's skeptical observations on the necessity and usefulness of the concept, Fleischer himself abandoned the category in a later edition of his text on word-formation in German (cf. Fleischer and Barz 1995: 27 ff.).

However, the class of phenomena recognized by the term affixoid has not been completely abandoned. On the contrary, its value has recently been recognized in a diachronic sense as proof of the continuum between composition and derivation (Stevens 2005; Lachachi 2008).[41] In a synchronic sense, however, it seems appropriate to observe the existence – sometimes taken as a further defining criterion of the class of affixoids (Kühnhold *et al.* 1978: 427- 518) – of a paradigmatic solidarity that in rather homogeneous semantic sets exists between different attested forms, despite the different meanings of their lexical sources. Consider, for instance, the common semantics of formations involving among others *-reich* (lit. 'rich'), *-voll* (lit. 'full'), *-stark* (lit. 'strong') and *-schwer* (lit. 'heavy'), which can be paraphrased as 'having a lot of X', where X stands for the element that occupies the first position. It should be noted that one of the functions of the suffix *-ig* roughly corresponds to this semantics.[42]

The affixoid is thus intended as an element that is no longer a word, but is not yet an affix either. The concept rests on the observation of how some units in word-formation – such as *frei* 'free' in *atomwaffenfrei* 'nuclear-free, devoid of nuclear weapons' – have a specific meaning that is often more salient than is oth-

[41] Fleischer and Barz (2012: 61) state that the concept of affixoid has proven useful in diachronic studies which focus on showing the graduality of the development of affixes from roots. However, in the synchronic perspective they take in their work on German word-formation, they consider elements like *riesen-* 'terrific, giant' (as in *Riesenspaß* 'terrific fun'), *voll-* '-ful', *-reich* 'rich' (as in *kontrastreich* 'rich in contrast'), and *-arm* 'poor' (as in *fettarm* 'low-fat') as bound roots with a specific interpretation (*lesartspezifisch gebundene Stämme*).

[42] On the presence of duplicates of similar function in word-formation compare Muthmann (1994) and Lachachi (2008), in addition to the semantic classifications presented in Kühnhold *et al.* 1978.

erwise granted to affixes. If this tends to grant these elements the status of words, other aspects – such as the tendency to distinguish between free and bound use of these elements – tend to favor their affixal status instead – compare *Werk* 'work' and its use conveying a collective nuance in *Astwerk* 'branches', *Buschwerk* 'intricate wood', *Laubwerk* 'foliage'. This uncertainty between the two levels has led to the proposal of the intermediate class of affixoids. Introducing this class, however, complicates rather than simplifies the classification problems of word-formation processes, as it creates two boundaries (one between affixes and affixoids, and the other between affixes and composition elements) where previously only one existed. For this reason, the class of affixoids is today frequently rejected as an independent category of word-formation, and there is a tendency to consider the border elements as belonging to one of the two traditional classes of affixes or compound elements, even if they lack some of the characteristic traits of such classes (Donalies 2002: 25; Lohde 2006: 15-16).

Other scholars, however, maintain the concept of affixoid in their studies. Below, a recent definition of affixoid is put forward, summarizing the main features that have been proposed to define this class.

«Semi-affixes are affixes that produce series of formations and act on existing lexemes, blurring their meaning. They correspond to lexemes of which they are homonyms. They cannot realize their affixal meaning independently, they are therefore related morphemes.» (Römer 2006: 198; my translation)[43]

Elsen (2011: 75-77) supports the usefulness of the class of affixoids by observing how they form open classes of elements on the border between compounding and derivation, which are systematically similar and relatively coherent. The traits that he indicates to define the class are divided into traits in common with the affixes and traits in common with the elements of compounding. The features in common with the affixes are:

i. the status of bound elements,
ii. the fixed position (bivalent trait through which it is possible to distinguish between prefixoids and suffixoids),
iii. the ability to create "automatic" series of formations,

[43] «Halbaffixe sind reihenbildende Affixe mit einer verallgemeinernden Bedeutung zur semantischen Nuancierung von vorhandenen Lexemen. Sie haben homonyme Lexementsprechungen. Sie können ihre Affixbedeutung nicht allein realisieren, und sind deshalb gebundene Morpheme.» (Römer 2006: 198)

iv. the impossibility of being directly connected to affixes,
v. the lack of lexical meaning.

Features in common with compounding are:

i. the presence of a free lexical correspondent,
ii. a greater phonological content and
iii. the possibility, at least for prefixoids, to bear the stress (on the contrary, most affixes are monosyllabic or phonologically reduced and unstressed).

After this survey of affixoids in German, we can now consider the nature of *selbst-*, and posit the question whether it can be interpreted as a compound element or as an affixoid.

First of all, the $_{CC}$marker *selbst-* corresponds to a syntactically free element which occurs in the following functions, and can be recognized as belonging to different word classes. This element finds a formal correspondence with:

a. the noun *(das) Selbst* '(the) self, (the) ego', which is diachronically a result of the substantivization of:
b. *selbst*, a focus particle (*Fokuspartikel*) appositive (*der Präsident selbst*, *er selbst*), sometimes involved in reflexive constructions (*sich selbst*) with the function of restricting the binding domain of the reflexive *sich*, facilitating the interpretation of coreference with a non-prototypical antecedent (*Die Eltern überließen die Kinder*$_i$ *sich*$_i$ *selbst* 'Parents left their children to fend for themselves') (Siemund 2007: 718-719); or alternatively as the substantivization of
c. *selbst*, a modifier of the verb phrase that can be paraphrased with 'independently' or 'without external help', sometimes also characterizing the agent as the beneficiary of the action (*Paul hat das Haus selbst gebaut* 'Paul built the house (for) himself') (ibidem).

According to many common definitions of compounding (see for example Guevara and Scalise 2008, quoted below in 2.2.8), the elements involved in an operation of compounding should belong to major syntactic categories, and in the case of German *selbst-*, similarly to *auto-*, we can identify a noun as a lexeme

corresponding to the element used in the compound. In the case of German, the noun *das Selbst* certainly has a closer connection than *auto-* with the elements involved in syntactic constructions related to reflexivity. However, it must be said that the lexical features of the name *das Selbst*, defined in the DUDEN *das seiner selbst bewusste Ich* 'The ego aware of itself', are not useful in the interpretation of German $_{CC}$formations.[44] Concerning the other two syntactically free elements proposed above, from a purely formal point of view, they are almost equivalent to the noun as candidates for the lexical sources of *selbst-* in bound use. However, from the point of view of their functions, if we consider, as we have done so far, the reflexive function of *selbst-*, neither of them performs it in free use, since this is essentially the prerogative of the reflexive pronoun *sich*.

A different conclusion can be drawn if we consider the functions of *selbst* as a focus particle and of the sentence adverb *selbst* – so essentially of the intensifier *selbst*, see chapter 1, section 2 – and compare them with the functions of *selbst-* as a $_{CC}$marker. In that case, the correspondence is rather high.

It is worth discussing here how compounds with *selbst-* are treated in the German morphological literature. In Fleischer and Barz (1995), *selbst-* is considered in the characterization of the left constituents of nominal and adjectival compounds. In both cases, *selbst-* is treated as a pronoun alongside more prototypically pronominal elements such as *ich* 'I', *wir* 'we' (*Ichform* 'narration in first person', *Wirbewusstsein* 'group consciousness'), or the possessive *mein* 'my' and *dein* 'your (SG)' (*der Dein-Tag und der Mein-Tag* 'the your-day and the my-day'; example derived from Ortner and Ortner 1984). In general, compounding with pronominal elements is judged to be rare, and limited to some subclasses of elements or to isolated cases (Fleischer and Barz 1995: 113). However, in its bound form, *selbst-* plays a particularly important role as first element in both nominal and adjectival compounds (Fleischer and Barz 1995: 114, 148). According to Fleischer and Barz, the development of the noun *Selbst*, attested ever since the seventeenth century, has been beneficial to the use of the "pronoun" *selbst* in compounding.

As mentioned above, however, the coreference function performed by *selbst-* in

[44] We can nonetheless find some examples in which *selbst-* in a compound should certainly be interpreted as the noun *(das) Selbst*. See, for example, the compounds *Selbstauflösung* 'dissolution of the self', or *Selbstbefreiung* 'liberation of the self', but also 'self-liberation'. It must be remarked that these examples seem to be very rare.

compounding is not shared by the free words that can be formally linked to it. This mismatch could bring *selbst-* closer to the class of affixoids.

Accepting the validity of the concept of affixoid and considering *selbst-* in the light of the features identified by Elsen (2011), as mentioned above, we might include it in the category of prefixoids since, like affixes:

i. its reflexive meaning is limited to its usage as a bound element,
ii. it has a fixed position as a left element in compounds where it occurs,
iii. it forms series of formations (especially with deverbal derivatives as base words),
iv. [this feature does not apply, see below]
v. it has no lexical meaning (i.e. it rarely has the meaning of the noun *das Selbst* in compounds),

However, like compounds units:

i. it has free correspondents (of a major word class, in the case of *(das) Selbst*),
ii. although monosyllabic, it is composed of an abundant series of phonemes (six),
iii. it carries the main accent of the compound (*Selbstachtung* 'self-esteem' ['zɛlpstaxtʊŋ])
iv. unlike affixes, *selbst-/Selbst* can be combined with affixes (*selbstisch* 'self-ish', *slebstlos* 'self-less').

We will consider *selbst-* and its complex network of relations again in 2.3.4 below.

2.2.7 Borderline units: German separable preverbs

We now turn to the second borderline phenomenon in German word-formation mentioned above in 2.2.5, separable complex verbs (SCVs; in German *trennbare Verben* 'separable verbs', or *Partikelverben* 'particle verbs'), which constitute a borderline case between derivation and syntax, and represent a widely discussed phenomenon which is especially prominent in German (Eisenberg 2012: section 7.1.2) and Dutch (Los *et al.* 2012).[45]

[45] See, however, Los *et al.* (2012: 7-11) for a survey of the phenomenon beyond West-Germanic languages.

SCVs are characterised by the fact that they constitute a morphological unit only in some of the syntactic configurations in which they occur.

(25) a. *an·kleben* 'to stick up'
b. *Sie muss den Zettel an·kleben* 'She must stick up the leaflet'
c. *Ich weiß, dass sie heute den Zettel an·klebt* 'I know that today she is sticking up the leaflet'
d. *Sie klebt den Zettel an* 'She sticks up the leaflet' (Eisenberg 2012: 255)
e. *Es ist schön, den Zettel an·zu·kleben* 'It's nice to stick up the leaflet'
f. *Sie hat den Zettel an·ge·klebt* 'She has stuck up the leaflet'

The verb *ankleben* in (25) illustrates the main properties of SCVs in German. They occur as syntactical and morphological units only in the infinitive, such as when they are governed by a modal (25).a-b, or in finite forms, when they occur in a subordinate clause (25).c. However, in a main clause, as in (25).d, the base verb is in second position, while the preverb, or particle, is in final position, determining that the lexeme *ankleben* no longer constitutes a unit either syntactically or morphologically. Interestingly, the morphological unit of preverb and base verb can also be interrupted in the case of a *zu*-infinitive, as in (25).e, or when the SCV occurs in the past participle, as in (25).f. However, here both preverb and verb, in this order, occupy the syntactic position at the end of the clause – compare the prefixed verb *ver·kleben* 'stick together', whose past participle is *ver·klebt* (**ge·ver·klebt*, **ver·ge·klebt*) and whose *zu*-infinitive is *zu ver·kleben* (**ver·zu·kleben*). Similar properties can be recognized in Dutch SCVs (Los *et al.* 2012: 60).

Despite the issues concerning their syntactic and morphological unity, SCVs also show word-like features. They constitute the input of word-formation operations such as compounding (26).b and derivation (26).c; they can operate a change to the valency of the base verbs, see (27).a, similar to that of prefixed verbs (27).b; they constitute a unit from a semantic point of view, often displaying an idiosyncratic meaning (Los *et al.* 2012: 55 ff.).

(26) a. *voor-leez-en*
for-read-INF
'to read to someone, to read out (a notice)'

b. *voor-lees-boek*
for-read-book
'a book to read to someone'
c. *voor-lez-er*
for-read-er
'one who reads to someone else'

(27)

a.	*SCV*	*base verb*
	de schoenen in-lopen	*(* de schoenen) lopen*
	the shoes in-walk	(the shoes) walk
	'to break in the shoes'	'to walk (the shoes)'
b.	*prefixed verb*	*base verb*
	de straat be-wandelen	*(* de straat) wandelen*
	the street PREF-walk	(the street) walk
	'to walk the street'	'to walk (the street)'

The examples provided in (26) and (27) concern Dutch, but similar ones can be found to illustrate the same features in German. Compare the Dutch examples with the German SCV *vor·lesen* 'read aloud, read out', and the compound *Vor·lese·buch* 'book to be read aloud (especially to children)', or the agent noun *Vor·leser* 'lecturer', all structured as the Dutch examples. Concerning valency changes, compare the sentence *Sie klebt den Zettel an die Wand* 'She sticks the leaflet on the wall', with the base verb *kleben* 'stick' and the examples in (25) above (for further examples see Eisenberg 2012: 252 ff.).

The last feature mentioned above, i.e. SCVs as semantic units, is of particular relevance because it draws our attention to the place of SCVs with respect to their inclusion in the lexicon. It is usually considered that complex words are stored in the lexicon, while syntactic constructions are generated online. However, in the case of SCVs, on the one hand, it is evident that their two constituent parts have different distributional constraints in different constructions, and in most of them occur in different positions within the sentence. On the other hand, they must be stored in the lexicon because of their word-like behavior, and especially because their meaning cannot be deduced only from the semantic contribution of their parts.

This idiosyncrasy can be exemplified by the meaning of SCVs with *op* in Dutch (see Los *et al.* 2012). In the SCVs, *op-* can be interpreted as meaning 'physically/

cognitively/perceptually accessible'. For example, *de boek opzoeken* can be paraphrased as 'to cause books (*boeken*) to become accessible by searching (*zoeken*)'. Los *et al.* (2012) suggest a series of steps of semantic extension by which one can connect the meaning of *op-* in SCVs with the spatial meaning of the preposition *op* 'up(ward), on high', but the important point to be made here is that the meaning 'accessible', expressed by *op-* in SCVs, is unavailable in other contexts, i.e. it is construction-specific (Los *et al.* 2012: 17).

Los *et al.* (2012) also discuss the grammaticalization cline in (28), by which a preverb is progressively integrated into a word, losing its head status.

(28) projecting preverb > optionally projecting preverb > non-projecting preverb > prefix > (zero)

SCVs in Dutch (and German) belong, according to Los *et al.* (2012), to the stage in which preverbs optionally project:

> «Our core hypothesis for the grammatical status of the particle generally is that the particle as a lexical head projects optionally: the default option is that the particle does not project […]. If required, the particle may project a phrase. The hypothesis of optional projection substantively accounts for the ambivalent morphosyntactic behaviour of particles.» (Los *et al.* 2012: 11)

It must be noted that not only prepositions can act as particles in SCVs, but also adverbs, adjectives, verbs and nouns. The *Substantivischen Verbpartikeln* ('substantive verb particles') have a rather particular behavior in German, since – despite being considered among the particles occurring in SCVs – they are actually inseparable from the verb, at least as far as they are produced by back-formation from a compound noun, as in the case of *not·landen* 'to perform an emergency landing' ← *Not·landung* 'emergency landing', *ehe·brechen* 'to commit adultery' ← *Ehe·brechen* 'adultery', or a compound participle (used as adjective) as in *schutz·impfen* 'to inoculate with a vaccine, to vaccine' ← *schutz·ge·impft* 'inoculated with a vaccine' (Fleischer and Barz 2012: 427-440).[46] These SCVs occur

[46] It is worth noting that *ehebrechen* and *schutzimpfen* might well have been back-formed from the synthetic compound *Ehebrecher* 'adulterer, or from the nominal compound *Schutzimpfung* 'vaccination' respectively.

in fact only in the infinitive (*notlanden*) – where they constitute a morphological unit – in the past participle (*not·ge·landet*), or in the *zu*-infinitive (*not·zu·landen*) – where the nominal preverb is separated from the base verb – and possibly in the finite forms only when they occur in final position in a subordinate clause. See the final position of *notlandet* 'perform an emergency landing' in this relative clause extracted from deTenTen13: *eines extraterrestrischen Volkes, das von seiner Welt vertrieben wird und auf der Erde* ***notlandet*** 'an extraterrestrial nation which has been expelled from its planet and ends up landing on Earth'.[47]

German nominal particle verbs appear at the crossroads of several morphological phenomena: unproductive NV compounding, fuelled by back-formation, joined with a syntactic behaviour close to that of more usual SCVs lying on the grammaticalization cline by which preverbs develop into prefixes.

Also, the case of SCVs shows how borderlines in morphology run not just between two domains; often even more domains come in contact with one another at the same time.

2.2.8 Compounds and the lexicon-syntax boundary

Although it is normally believed that the products of word-formation are stored in the lexicon, and that the products of syntax are formed online instead, SCVs suggest that the issue of the boundary between word-formation and syntax is not so easy to dismiss.

Gaeta and Ricca (2009) considered this dichotomy, especially concerning compounds, and have shown that the relationship between the lexicon and the processes underlying the formation of phrases and compounds cannot be structured solely on the basis of this dichotomy. They observed that there are linguistic structures which, though clearly produced syntactically, are nonetheless stored in the lexicon. On the other hand, there are compounds that are not stored in the lexicon, but which are clearly identifiable as products of morphological operations.

In order to distinguish between phrases and compounds, Gaeta and Ricca (2009) have therefore added another feature to the traditional dichotomy [± lex-

[47] Fleischer and Barz (2012: 440) add that there is a growing tendency for back-formed nominal particle verbs to also occur in main clauses as non-separable complex verbs. See, for example, the position in this main clause, extracted from deTenTen13: *Irgendwann* ***notlandet*** *der Autor Antoine in einer Wüste* 'The author Antoine performed an emergency landing somewhere in the desert'.

ical], namely [± morphological], which indicates that a morphological operation is at play in the formation of a linguistic structure. Crossing these two dimensions generates the typology in (29) (Gaeta and Ricca 2009: 38).

(29) a. [+ morphological], [+ lexical]
b. [+ morphological], [– lexical]
c. [– morphological], [+ lexical]
d. [– morphological], [– lexical]

In this typology, (29).a constitutes the prototypical case of compounding, while (29).d constitutes the prototypical case of phrase. (29).c configures the cases of lexicalized phrases, such as Italian *luna di miele* 'honeymoon', which denote unitary concepts, have an idiomatic meaning that is difficult to infer from that of their components – we cannot infer form *luna* and *miele* the meaning of 'holiday following marriage' – cannot be interrupted by any type of linguistic material, and yet are structured as phrases – *luna* is modified by the prepositional phrase *di miele* in the same way as any other noun might be: *farina di riso* 'rice flour'. The case of phrases which are stored in the lexicon and are strongly idiomatic, although marginal with regard to prototypical cases of the products of syntax and compounding, has not escaped the attention of Di Sciullo and Williams (1987), who recognized them as listemes. A similar perspective is present in Jackendoff (1997), who proposes a concept of lexicon which includes both elements inserted in syntagmatic structure at X0 (grammatical words belonging to the categories of N, V, A, P), and structures at a level higher than X0 that have an idiomatic nature.

On the other hand, the type listed in (29).b is a case that has so far received little consideration in the discussion on compounding. Many definitions of compounding include among their criteria the status of lexeme/listeme (for example, the definition in Bauer 2001a: 695),[48] an inclusion that can be explained by the fact that compounding is a process of word-formation, understood by many scholars as lexeme formation (Aronoff 1994: 13-16). So based on the idea that a lexeme is an element stored in the lexicon, Gaeta and Ricca define (29).b as "non-lexeme-forming compounding". They also provide various examples of this

[48] «We can now define a compound as a lexical unit made up of two or more elements, each of which can function as a lexeme independent of the other(s) in other contexts, and which shows some phonological and/or grammatical isolation from normal syntactic usage.» (Bauer 2001a: 695)

type of formations, both from the domain of derivation as well as the domain of compounding. It should be stressed that it is possible to identify highly productive processes that result in words which are not stored in the lexicon, or even in those that cannot be stored: for example, nonce-formations containing proper names or formations strictly linked to the context of the utterance, for which it is difficult to imagine that they will ever be permanently included in the lexicon.[49] Italian examples can be the deanthroponymic derivatives in *-iano* (*gli obbiettivi più probabili della rappresaglia gheddafiana* 'the most probable targets of Gaddafi's retaliation' *La Repubblica*) and VN compounds (*quel terzo posto* ***acchiappa-Uefa*** 'that Uefa-catching third place, *colpo di testa* ***fissa-risultato*** 'score-setting header').[50]

The focus of interest of Gaeta and Ricca (2009) is to bring clarity particularly to the boundary between compounding and syntax. For this purpose, they indicate and discuss certain criteria for the definition of compounds on a formal basis and for their distinction from the products of syntax.

Identifying the procedures that define the [+ morphological] feature of the typology in (29) is a difficult task to achieve on a cross-linguistic basis. The same holds for identifying cross-linguistically valid phonological criteria which are able to distinguish compounds from phrases on the one hand, and on the other, to assimilate compounds and phonological words (see Arnaud 2004a: 329-332).

Despite the possible issues, Gaeta and Ricca propose three criteria which, although very general and subject to some limitations, can help to distinguish compounds and phrases on a formal basis. The first criterion is that of non-interruptibility:[51]

[49] It is worth noting that the lexicalization and idiomatization of lexemes derived from proper names is not impossible, but it usually follows a process of antonomasia, and consequently, that of abstraction from the context of the first use. See e.g. the English verb *bowdlerize* «to remove the parts of a book, play, etc. that you think are likely to shock or offend people […] Named after Dr. Thomas Bowdler, who in 1818 produced a version of Shakespeare from which he had taken out all material which he considered not suitable for family use.» (Oxford Advanced Learners' Dictionary 2005)

[50] AA compounds display a similar tendency at forming nonce-words as suggested by Grossmann and Rainer (2009: 90-91): *inciucio berluscon-dalemiano* 'under-the-counter deal between Berlusconi and D'Alema', *meccanismo utilitar-affaristico* 'utilitarian-profiteering mechanism'.

[51] Gaeta and Ricca (2009) consider in this regard the case of German separable complex verbs such as *radfahren* 'to ride a bicycle' as a case that can be referred to compounding. Not all scholars agree with this interpretation, since they consider the criterion of non-interruptibility in a more restrictive way. Blom (2005) and Booij (2010: chap. 5) consider the similar case of Dutch separable complex verbs (*groot brengen* 'to grow', literally 'to bring big') as an example of constructional idiom, i.e. a complex syntactic structure with a non-compositional meaning and therefore stored in the lexicon (se also above 2.2.6).

> «[T]he compound must consist of one continuous phonological string, which cannot be interrupted by any intervening (non-inflectional) linguistic material. […] [S]tructural adjacency refers to the fact that the lexical items involved in a compound have to be close to each other at least at one configurational level.» (Gaeta and Ricca 2009: 43)

This is not a sufficient condition, due to the presence of non-interruptible lexicalized phrases, but still holds for defining compounds.

The second proposed criterion is present in most definitions of compounds:

> «[T]he compound must be made up of at least two lexical morphemes» (Gaeta and Ricca 2009: 44)

This criterion makes it possible to exclude prefixed words and sequences of auxiliary and verb that could overcome the criterion of non-interruptibility from the set of compounds. As the authors comment in one footnote, the exclusion of PN formations is less obvious: for some of them, it is easy to recognize a process of univerbation, i.e. the process by which a syntactic sequence stabilizes to the point of being stored as an indivisible and unanalysable unit (e.g. *insomma* 'in short, in a word' lit. 'in sum', *adagio* 'slowly' lit. 'at ease'). For other PN formations, a certain regularity of the pattern on which they are based, or the presence of a preposition that can also act as an adverb, configures a borderline area between derivation and composition and between lexical and grammatical/functional elements.

The last criterion proposed by Gaeta and Ricca (2009) concerns the elements that have probably generated most confusion in defining what should be considered a compound and what should not:

> «[T]he compound must not be structured by means of any functional word which codifies inner grammatical relationships between its components» (Gaeta and Ricca 2009: 44)

The functional words in question are prepositions, articles or conjunctions, which in some languages provide the chunks of grammatical information necessary to define the internal relationships of the phrases. As can also be seen from

Bauer's (2001a: 699, 702-705) discussion, in some traditions of morphological description these elements are considered as strategies for giving formal cohesion to compounds – consider, for example, the category of French *composés prépositionnels* (*chemin de fer* 'railway', lit. 'iron path'). In other languages, these syntactic devices have shifted to become linking elements – see, for example, Danish coordinative compounds: *smør·e·brød* 'butter-LINK-bread', where the marker *-e-* develops from the phonological reduction of the coordinative conjunction *og* 'and'.

This aspect of the definition of compounding proposed by Gaeta and Ricca (2009) is in line with the definition provided by, for example, Guevara and Scalise (2008):

> «The inner essence of a compound can be captured (in the prototypical case) with the following rough schema, where X, Y and Z represent major lexical categories, and ℜ represents an implicit relationship between the constituents (a relationship not spelled out by any lexical item): [X ℜ Y] Z» (Guevara and Scalise 2008: 107)

If French *composés prépositionnels* are clearly a case of syntactic structure in spite of the high degree of lexicalization and idiomaticity, the case of Danish coordinative compounds can now be considered as a regular morphological operation, also because in synchrony it is not recognizable as a syntactic operation. On the contrary, if we take Italian *pomodoro* 'tomato', although it is probably unanalysable in synchrony as the phrase *pomo d'oro* 'golden pome/apple', «at no point of the lexicalization process from *pomo d'oro* to *pomodoro* it seems legitimate to speak of a compound in our morphological sense, because these items are in no way the output of a morphological rule» (Gaeta and Ricca 2009: 44).

2.3 Construction Morphology

The overview of borderline phenomena of the previous section has helped us to introduce $_{CC}$markers – at least Italian *auto-* and German *selbst-* – from the perspective of word-formation. On the other hand, that overview has showed the weaknesses of morpheme-based approaches to word-formation and of a rigid demarcation of the boundaries between word-formation and adjacent domains (inflection and syntax), but also, internally, between derivation and compounding. In this section, we will finally introduce Construction Morphology (henceforth CxM), a theoretical framework developed in the past two decades by Geert Booij, partly rooted in the tradition of Lexicalism, and partly converging on the theoretical assumptions of Construction Grammar (henceforth CxG), especially as far as the continuum between lexicon and grammar is concerned.

In this section, we will provide a profile of CxM by surveying the main theoretical assumptions (2.3.1), the theoretical devices that are used to relate constructions with each other (2.3.2), and the way CxM treats derivation, compounding and relevant borderline phenomena (2.3.3). Finally, we will give an account of Italian and German $_{CC}$markers from a CxM perspective (2.3.4).

2.3.1 Theoretical assumptions of Construction Morphology

As mentioned above, Construction Morphology is a morphological theory that is consistent with the general framework of Construction Grammar (Kay and Fillmore 1999; Goldberg 1995, 2006). A crucial concept of CxG is that of construction defined as follows:

> «Constructions are stored pairings of form and function, including morphemes, words, idioms, partially lexically filled and fully general linguistic patterns.» (Goldberg 2003: 219)

This definition by Adele Goldberg also fits well into CxM, making it a «sign-based theory of morphology whose building blocks are the constructions» (Masini and Audring 2019: 366). CxM is, however, a word-based theory, so the minimal sign is not the morpheme as in Goldberg's definition, but rather the word (Booij 2010: 15). On the other hand, CxM shares with Goldberg's CxG the idea

that constructions are «**learned** pairings of form with semantic or discourse function» (Goldberg 2006: 5; my emphasis), consequently characterizing CxM as usage-based theory.

This calls for a more detailed account of how constructions interact with each other and with the lexicon and grammar in CxM. From this point of view, it is worth to consider the link between CxM and Jackendoff's Parallel Architecture.[52]

In the Parallel Architecture, «linguistic structure is determined by three independent systems of representation – phonology, syntax and semantics – and by the linkages among them» (Jackendoff and Audring 2019: 391), as can be schematically represented in Figure 1.

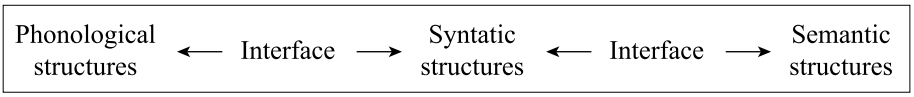


Figure 1. The Parallel Architecture

This assumption is adopted in CxM, so that both simple (30).a and complex words (30).b are constructions that can be modeled through a tripartite parallel structure.

(30) a. $<([\text{dɒg}])_{\omega\text{-i}} \leftrightarrow N_i \leftrightarrow \text{DOG}_i>$
b. $<(([\text{beɪ})_\sigma\ (\text{k}]_j[\text{ər}]_k\ )_\sigma\ )_{\omega\text{-i}} \leftrightarrow [V_j\ \text{Suff}_k]_{Ni} \leftrightarrow [\text{one who BAKE}_j]_i>$

The representation in (30).b shows how the different parts of the structure can relate to each other via co-indexation.[53] It is also evident that syllabic structure and morpho-syntactic structure are not isomorphous. Also, it is worth noting that the co-indexation of the affix *-er* is needed only to correlate a piece of phonological structure with the morpho-syntactic structure, but does not appear in the semantic structure. As a matter of fact, affixes are not stored in the lexicon, and do not exist independently of the construction in which they occur (see further below).

Considering the case of a family of syntactic constructions – the NPN construc-

[52] See Jackendoff (1997, 2002) and Jackendoff and Audring (2019) for an overview focused of morphology. Booij and Audring (2015) provide a detailed assessment of the points of contact between the Parallel Architecture and Construction Morphology.

[53] Examples (30) and (31) are adapted from Booij (2010: 7-8), following the formalizations used in Booij and Audring (2015: 279) to represent the morphological construction $[\text{N-}\textit{hood}]_N$.

tion, e.g. N *by* N (*day by day*) or N *after* N (*picture after picture*) – Jackendoff (2008) suggests three important features of a constructional approach (Jackendoff 2008: 15-16; emphasis in the original):

i. «it allows pieces of syntactic structure to be listed in the lexicon with associated meanings, just as individual words are; these are the MEANINGFUL CONSTRUCTIONS of the language»;
ii. «construction grammar makes no principled distinction between words and rules: a lexical entry is more wordlike to the extent that it is fully specified, and more rulelike to the extent that it contains variables that have to be filled by other items in the sentence»;
iii. «lexical entries are arranged in an inheritance hierarchy, so that commonalities or redundancies among words and constructions can be captured by entries at a higher level in the hierarchy. That is, construction grammar gives up the assumption, inherited from structuralist linguistics, that the grammar captures all regularities, and the lexicon is simply a nonredundant list of exceptions, totally distinct from the grammar».

Leaving point i. aside for a moment, let us discuss the second feature of constructions mentioned above. Since complex words are constructions just like phrases, they can, as phrases, be either fully specified or contain variables. The series of formations similar to *baker* in (30).b above, can thus be represented as in (31) in the form of a construction with an open slot that can be filled with a lexeme, in the same way as a word-formation rule.

(31) $<([\ ]_j\ [ər]_k)_{ω\text{-}i} \leftrightarrow [V_j\ Suff_k]_{Ni} \leftrightarrow [\text{one who PRED}_j]_i>$

We must, however, point out that in CxM constructions are represented as schemas and not as rules. The difference is that, while word-formations rules, such as those proposed in Aronoff (1976), are essentially input-oriented, connecting an input to an output via one or more operations, schemas are output-oriented. They constitute «static generalizations over a set of fully specified items» (Masini and Audring 2019: 369), and are thus essentially declarative, while «word-formation rules are procedural and imply productivity by default» (*ibidem*). The last difference between rules and schemas lies in their position-

ing with respect to the borderline between the grammar and the lexicon. This leads us back to the features of constructions suggested by Jackendoff (2008) and listed above: syntactic constructions can be listed in the lexicon, but beyond that, CxG refuses the contrast between grammar as the list of regularities and the lexicon as the list of idiosyncrasies, by refusing the distinction between lexicon and grammar. So does CxM, and it is here that this last distinction between rules and schemas lies: while rules lean on the idea that they belong to grammar and their products are then stored in the lexicon, CxM simply does not make such a distinction. Just like CxG, CxM advocates for a continuum between lexicon and grammar, and a single repository for both fully specified constructions and morphological schemas, i.e. the constructicon (Masini and Audring 2019: 370).

A final point about constructions concerns their mutual relations. The constructicon is not simply an unstructured list of constructions of varying complexity, but a hierarchy (see again Jackendoff 2008: 15-16, quoted above). This leads us to the survey of the theoretical devices that operate in CxM.

2.3.2 Devices of Construction Morphology: relations between constructions

The first device that we shall consider is default inheritance:

> «[T]he specification of a word for a particular property is inherited from the dominant node, unless the actual lexical entry has another specification for the property. […] The effect of default inheritance is that information on a higher node may be superseded by information concerning the relevant property on a lower node. Hence, this system of computing properties is non-monotonic. This means that not all information that is derived from higher nodes is necessarily preserved.» (Booij 2010: 27-28)

Default inheritance is a very powerful device when dealing with idiosyncrasies at the level of individual complex words. See, for example, Dutch *werk-baar,* for which the constraint that requires transitive verbs as bases for deverbal adjectives in *-baar* is overridden, since *werk* 'to work' is an intransitive verb. As Booij (2010) points out, the adjective *werkbaar* is a regular adjective, but displays «an exceptional sub-categorization (co-occurring with an intransitive verb), which overrules

the subcategorization for transitive verbs in the relevant schema» (Booij 2010: 27).

However, when operating at the level of constructions, default inheritance displays even more far-reaching consequences in interaction with subschemas – i.e. schemas linked to higher level schemas that are consequently characterized by a lower level of abstraction. In this context, default inheritance licenses the existence of subschemas which override some of the properties of their higher nodes, thus allowing for schemas that are both productive and "idiosyncratic" (see the case of affixoids in 2.3.3 below).

Furthermore, default inheritance and subschemas allow for the definition of patterns of polysemy of a single affix. This can be exemplified by the Dutch affix *-er*, whose interpretation changes, depending on the class and the properties of the base words to which it is connected (see Figure 2 below).

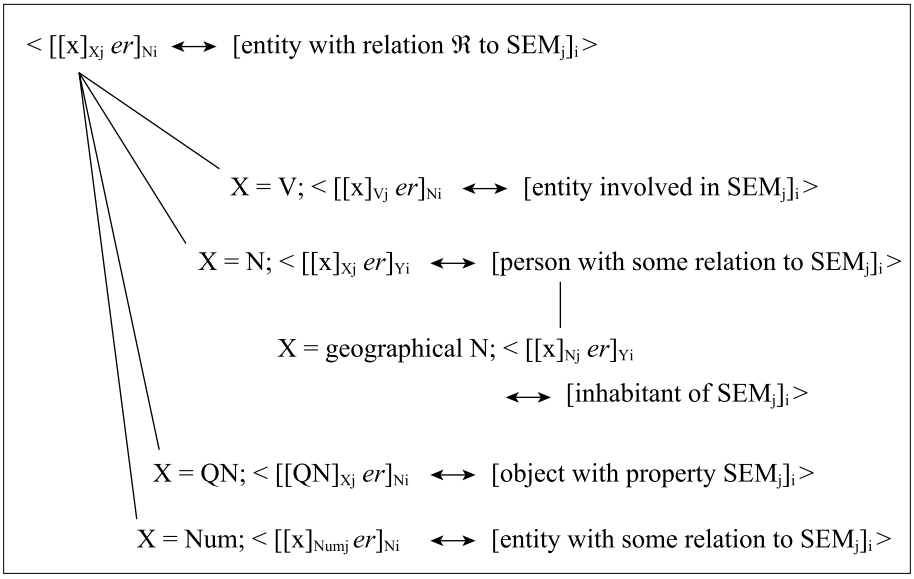


Figure 2. Subpatterns of the schema for Dutch nouns in *-er* (Booij 2010: 84; with modifications)

The generation of subschemas depends on several properties of schemas that Booij (2010: 41) derives from Rumelhart (1980): schemas have variables, can be embedded in other schemas, and represent knowledge at all levels of abstraction. The latter property is linked to default inheritance, since it allows for the existence of subschemas with specific properties overriding the properties of higher nodes.

The first two properties, variables and embedding, are at the core of CxM's schema unification (Booij 2010: 41 ff.).

In CxM, schema unification is the operation that allows the substitution of variables, i.e. open slots within a construction, with specific lexical items. The case of regular deverbal adjectives in *-baar*, such as *drinkbaar* 'drinkable', is an example of this type of unification (represented in Figure 3 below), but also constitutes an Instantiation Inheritance Link (I_I) which «connect[s] schematic constructions to progressively more specified constructions» (Masini and Audring 2019: 374).

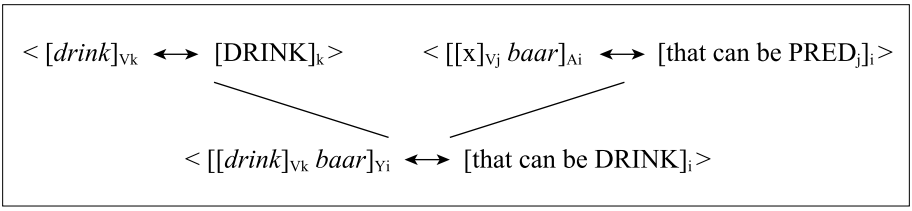


Figure 3. Schema unification: Instantiation Inheritance Link

Schema unification can also create semi-specified subschemas which are able to account for complex words based on unattested complex base words. Rule-based frameworks struggle with these cases, because rules are input-oriented, so an unattested step in the derivation chain creates well-known puzzles, such as the ones surrounding synthetic compounds.[54]

However, embedding through schema unification generates subschemas which can exist independently even of the productivity of an intermediate step. This is the case, for example, of NV compounds in Dutch, which display a very low productivity, but which are nonetheless productive when embedded in compounds, or as base nouns for *-er* formations (see Figure 4 below).

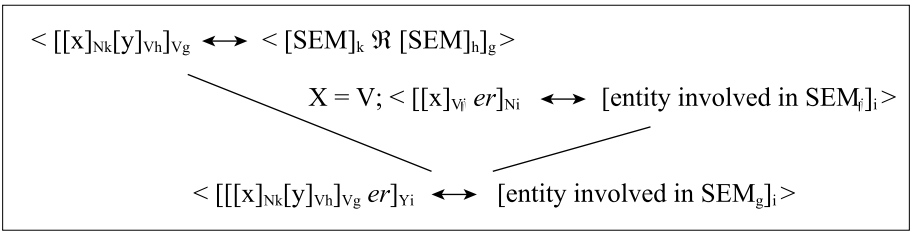


Figure 4. Schema unification: Embedded schemas

[54] See Gaeta 2010 for a survey of German synthetic compounds.

We can observe that the schema in Figure 4 is obtained by unifying one of the subschemas in Figure 2 – the one in which the Dutch suffix *-er* is connected to verbs – with the lowly productive schema of Dutch NV compounds. Isolated cases of NV compounds obtained by conversion ($[[[voet]_N[ball]_N]_N]_V$ 'to play football'), back-formation ($[[woord]_N[speel]_V]_V$ 'to play with words' ← *woord·spell·ing* 'word-play, pun'), or quasi-compounds ($[worst]_N\,[hap]_V$ 'sausage eat', compare the subordinate clause *...dat Jan worst hap-te* '…that John ate sausage')[55] do exist, but they are not sufficient to consider the schema productive. As a matter of fact, in the obtained schema there is no need for the specific NV compounds to be attested: see e.g. °*brand·bluss* → *brand·bluss·er* 'fire-estinguisher'. Consequently, we have a unified schema provided with a holistic property, productivity, which is not shared by either of the higher level schemas (Booij 2010: 47-50).

2.3.3 Derivation and compounding in Construction Morphology

We have already exemplified the treatment of protypical derivation and prototypical compounding by discussing the main theoretical assumptions and devices that are typical of CxM. To summarize, while compound schemas involve two or more lexical items provided with indices and subcategorised for certain properties, derivational schemas are characterized by a lexical item and an affix, i.e. a phonological string which does not exist independently of the construction, which means that it lacks an index, while the subcategorization properties are encoded on the morphological tier of the formal side of the construction.

(32) a. $< [[x]_{X\alpha k}[y]_{Y\beta i}]_{Y\beta j} \leftrightarrow [SEM_i \text{ with relation } \Re \text{ to } SEM_k]_j >$
b. $< [[x]_{X\alpha k}[y]_{Y\beta i}]_{X\alpha j} \leftrightarrow [SEM_k \text{ with relation } \Re \text{ to } SEM_i]_j >$
c. $< [[x]_{X\alpha k}[y]_{Y\beta i}]_{Z\gamma j} \leftrightarrow [SEM \text{ with relation } \Re \text{ to } [[SEM]_i\ \Re\ [SEM]_k]]_j >$
d. $< [x[y]_{Y\alpha i}]_{Y\alpha j} \leftrightarrow [SEM\ [SEM]_i]_j >$
e. $< [[x]_{X\alpha i}\ y]_{Z\beta j} \leftrightarrow \square\ [SEM\ [SEM]_i]_j >$

In (32), we show a series of schemas for right-headed compounds (a.), left-headed compounds (b.), exocentric compounds (c.), prototypical (i.e. category-maintaining) prefixation (d.) and prototypical (i.e. category changing) suffixation (e.).

[55] Compare the Dutch example with that of German SCVs with noun particles above in 2.2.5.

The differences in headedness are shown by the subscript uppercase Latin letters which represent the word class of the relevant lexical items, and by the subscript Greek letters which represent their subcategorization properties. Variables for both lexical items and affixes are represented by lowercase Latin letters, but, while lexical items are provided with a word class, subcategorization properties and an index (subscript lowercase Latin letter), affixes are not – e.g. in (32).d compare x, which represents a prefix, and y, representing a lexical item.

The characteristics of prototypical cases of compounding and derivation are also well captured by previous, rule-based and input-oriented frameworks. The advantages of CxM emerge in the treatment of problematic and borderline cases, such as the case of synthetic compounds considered in 2.3.2, the case of neoclassical combining forms, affixoids and SCVs considered in 2.2.4-2.2.5, 2.2.6 and 2.2.7, respectively. One of the advantages of CxM over previous frameworks is the possibility to create networks of schemas which establish lexical connectivity across the constructicon. Its function «is MOTIVATION, the reduction of arbitrariness of form and meaning in the daughter words [ ]. Obviously, being a motivated linguistic sign is a gradient property which correlates with the degree to which the properties of the schema are preserved in the daughters; hence, motivation is often partial» (Masini and Audring 2019: 374).[56] It is exactly partial motivation that is at work in the case of affixoids: while the form is motivated since it is inherited from a compound schema, the meaning is directly linked to the construction itself. Consider the case of the Dutch prefixoid *hoofd-* 'main' below.

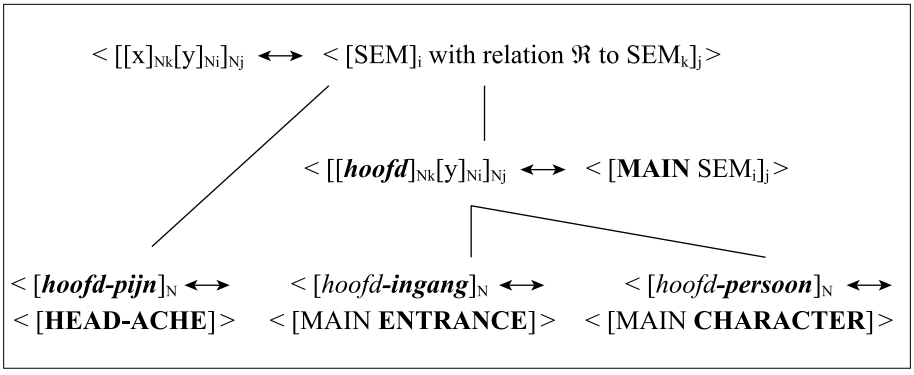


Figure 5. Hierarchy for Dutch *hoofd-* (Masini and Audring 2019: 386)

[56] See also Booij and Audring (2017) for the concept of (partial/multiple) motivation within CxM.

The schema for compounding motivates both the compound *hoofdpijn* 'headache' and the semi-specified schema *hoofd*-N 'main N', which is only partially motivated, since the contribution of *hoofd* 'head' in the schema is not compositional, and an unpredictable meaning 'main' is conveyed by the construction. In this sense, the schema displays features of both compounding and derivation. We can compare the case of Dutch *hood-* with German *Haupt-* 'main' (*Haupt·eingang* 'main entrance'; *Haupt·bahnhof* 'main train station'), where the only difference lies in the fact that *Haupt* – which is cognate to Dutch *hoofd* – has become obsolete in German and has been substituted by *Kopf* 'head'. This has lead German scholars to consider *Haupt-* as a prefix instead (Fleischer and Barz 2012: 257). Hüning and Booij (2014), however, stress the fact that in both cases CxM accounts for the constructions

> «by assuming a constructional schema with the first slot filled and a variable as the second element. The schema looks almost identical in both cases and the question whether we regard the first element as a noun, a prefixoid, or a prefix is not a question of principle. The differences concern mainly the position of the schema within the network of constructions: is it (still) associated with the more general schema for nominal compounds? Do language users (still) see the connection with the original noun?» (Hüning and Booij 2014: 591)

The difference between Dutch *hoofd-* and German *Haupt-* is thus a matter of different patterns of motivation in the hierarchical lexicon.

Until now, we have considered exclusively links that operate vertically, i.e. at different levels of abstraction within the hierarchy. An additional type of relation that can be exploited in CxM for accounting for problematic cases are paradigmatic relations. This kind of relation can be established between morphological schemas as in (33), but also between a morphological schema and a syntactic construction as in (34).

Schemas built on paradigmatic relations of the kind illustrated above are also called second order schemas (Masini and Booij 2015). In (33), their use is to account for a mismatch between meaning and form that has been treated as a case of affix substitution: a word like *alpinist* is linked with *alpinism*, rather than to the adjective *alpine*. Aronoff (1976) addressed this problem by positing a truncation rule, according to which e.g. the suffix *-ism* is deleted when the affix *-ist* is applied to the base *alpinism*. In CxM, the same phenomenon is instead treated with a paradigmatic relation that establishes co-indexation links between two schemas

– see (30).b above – so that the meaning of *-ism* formations has to be interpreted through the schema of the correspondent *-ist* formation.

(33) a. *alpinism* *alpinist* (Masini and Audring 2019: 380)
socialism *socialist*

b. $< [\text{x-}ism]_{Ni} \leftrightarrow [\text{SEM}]_i > \approx < [\text{x-}ist]_{Nj} \leftrightarrow [\text{PERSON involved in } [\text{SEM}]_i]_j >$

In (34), the same principle is applied to nominalizations derived from SCVs (see 2.2.5 above). In this case, the second order schema helps to model both formal and semantic aspects.

(34)

	verb	deverbal noun (Booij and Masini 2015: 62)
a.	*grijp* 'grab'	*greep* 'grip'
	in·grijp 'intervene'	*in·greep* 'intervention'
	mis·grijp 'slip up'	*mis·greep* 'slip-up'
b.	*zien* 'see'	*zicht* 'sight'
	aan·zien 'watch'	*aan·zicht* 'insight'
	op·zien 'supervise'	*op·zicht* 'supervision'

c. $< [\text{Part}_i\ V_j]_k \leftrightarrow \text{SEM}_k > \approx < [\text{Part}_i\ [V_j\text{-Nom}]_N]_{Nm} \leftrightarrow [\text{NOM } [\text{SEM}_k]]_m >$

On the one hand, it can be observed that the nominalizations of SCVs in (34) share the forms of the nominalization of the simple verb, which are formed through unproductive schemas involving stem allomorphy – as in (34).a: *grijp* → *greep* – or unproductive suffixes – as in (34).b: *zien* → *zicht*. On the other hand, while the form implies that these deverbal nouns are first derived from the simple verb, and only afterwards the preverb is attached to them, the meaning of the deverbal nouns linked to a SCV corresponds to the often idiosyncratic meaning of the SCV. This is again captured by coindexation across schemas, as can be seen in (34).c above.

2.3.4 Italian and German contrastive coreference markers: a Construction Morphology account

We can now apply the model of CxM to Italian and German $_{CC}$markers. We will see that Italian and German differ crucially in terms of the networks of constructions in which they are included. Italian constitutes a rather prototypical case of

derivation, while the German case implies a more complex network of relations, which in our opinion also involves a paradigmatic relation.

Figure 6 / a. illustrates the schema of compounds with *auto* – a "second generation" neoclassical combining form, clipping of *automobile* 'car'; see 2.2.4 above – which are structured according to the rules of neoclassical combining forms, making them right-headed, contrary to usual Italian compounds.[57] From a synchronic point of view, the formal link between *auto* 'car and the $_{CC}$marker *auto-* can be considered purely accidental, whereas diachronically, the two elements are linked in a direction that is very much opposite to the one that would be implied through schema unification – *auto* 'car' is a clipping obtained from *automobile* 'car'.

As discussed above in 2.2.4-2.2.5, despite its origin within the inventory of neoclassical combining forms, *auto-* is considered a prefix in the literature. Figure 6 / b. represents the schema of the $_{CC}$marker *auto-* as a semi-specified schema, subordinate to the general schema for Italian prefixation. It is worth noting that the formal sides of the schemas in a. and in b. differ only in the fact that in a. *auto* has an index, while in b. *auto-* doesn't: the latter is an affix and it is not listed in the lexicon independently of the construction in which it occurs. For the schema, we have chosen one of the possible functions of $_{CC}$markers – the one corresponding to the adverbial exclusive intensifier.

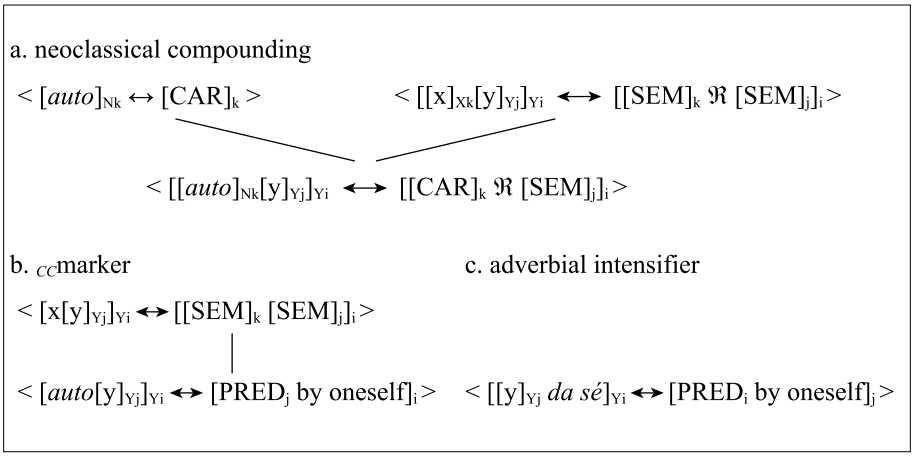


Figure 6. Italian compounds with *auto* 'car'. Italian $_{CC}$marker *auto-*. Italian adverbial exclusive intensifiers *da sé*.

[57] It is worth noting that right-headed compounds with *auto* 'car' can compete with left-headed ones: *demolizione auto* 'car scrappage' vs. *autodemolizione* 'id.'.

In Figure 6 / c., we have also included the construction with the Italian adverbial exclusive intensifier *da sé*, to show that there is a semantic correspondence between the two, but no formal similarity, which is why no link between the two can be established.

Quite a different situation is configured by the German $_{CC}$marker *selbst-*, whose network of relations is represented in Figure 7.

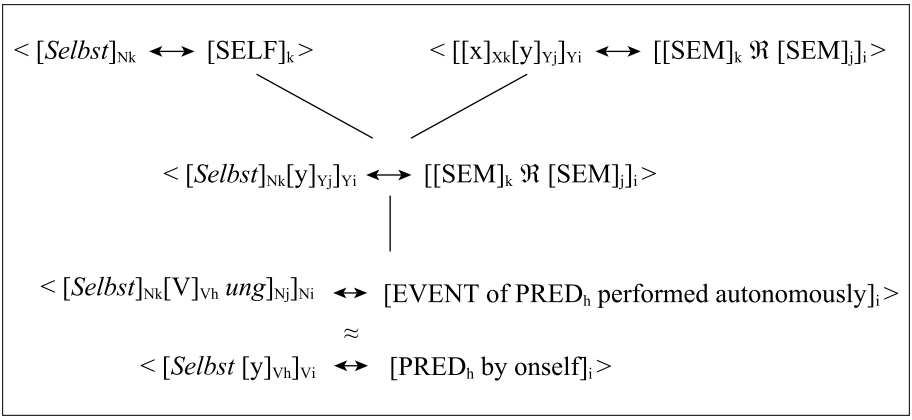


Figure 7. German $_{CC}$marker *selbst-*

The case of *selbst-* is closer to compounding, since the existence of the noun *(das) Selbst,* and the availability of a handful of compounds in which *(das) Selbst* contributes with its lexical meaning 'self, ego', justify considering the schema with *selbst-* as a $_{CC}$marker subordinate to the more general schema of German NN compounds. On the other hand, the meaning of the schema for the $_{CC}$marker deviates from the lexical contribution of the noun *(das) Selbst*, as can be seen by the absence of elements in the meaning side of the schema, coindexed with the first constituent [*Selbst*]$_{Nk}$. This makes the case of *Selbst-* rather similar and close to that of Dutch *hoofd-*, as seen above in 2.3.3 and in Figure 5. We have considered linking the schema for *selbst-* as a $_{CC}$marker under the schema of regular NN compounds with *(das) Selbst*, in order to stress the fact that the schema is as productive as NN compounds are, while compounds based on closed class elements like pronouns are far less productive. This reflects the observation by Fleischer and Barz (2012), already quoted above, that the coinage of *(das) Selbst* might have favored the productivity of *selbst-*.

A final important detail of the network in Figure 7 can be found in the paradigmatic relation of the schema of $_{CC}$formations with the construction of the adverbial exclusive intensifier *selbst*. Unlike Italian *auto-*, the intensifier *selbst* is both formally – it is a homophone of the noun *(das) Selbst* – and semantically related to the $_{CC}$marker, to the point that we can coindex across the paradigmatic relation, and motivate the meaning of $_{CC}$formations through the link with the construction containing the intensifier – and not, for example, with *sich* and the reflexive construction.

2.4 Conclusion

In this chapter, we have discussed the main theoretical concepts that will inform the discussion of the cross-linguistic data in Chapters 4-6.

We have surveyed phenomena at the boundaries between word-formation and adjacent domains within and outside morphology, but we have mainly considered the boundary between derivation and compounding.

Concerning this latter borderline, CxM has helped us in modelling the difference between the two main subdomains of word-formation as far as the network of motivation links in the hierarchical lexicon is concerned. The difference between derivation and compounding can be understood in terms of different degrees of motivation of the phonological material and semantic contribution of a morphological operation: if the formal and semantic modifications that a base word undergoes once a morphological operation is applied, display correspondence with independent lexemes, but also with higher, more abstract schemas in the lexicon, or whether a formal or semantic correspondence is established through paradigmatic relations with other constructions (either morphological or syntactic).

The correspondence with an independent lexeme can be completely absent, and a constructional schema can thus be completely conventional (prototypical derivation). In other cases, the correspondence with an independent lexeme can be complete, and the output of a constructional schema completely motivated (prototypical compounding).

The cases of Italian and German $_{CC}$markers have provided us with two quite different patterns of motivation, with Italian *auto-* very close to prototypical deriva-

tion, and German *selbst-* closer to compounding, but quite far from its prototype, sharing instead many features that have been recognized as typical of the disputed class of affixoids.

The analysis of Italian and German $_{CC}$markers will be used as a starting point for both gathering the data from the languages of the sample, and comparing the case of $_{CC}$markers in other European languages. In the next chapter, we will introduce the issues of the cross-linguistic study of word-formation, and we will also introduce and complete a profile of the languages included in the sample.

3

WORD-FORMATION CROSS-LINGUISTICALLY

In the previous chapter, we have surveyed the theoretical issues connected to the study of word-formation and its borderline phenomena. We have concluded the chapter by introducing Construction Morphology and its formalism, applying it to the case of $_{CC}$markers in Italian and German.

In this chapter, we will turn our attention to the study of word-formation cross-linguistically (3.1), and we will introduce the cross-linguistic sample – albeit areally limited to Europe – and the methodology used in this study (3.2).

3.1 Typology and the study of the lexicon

In this first section, we will discuss some issues connected to typology and the study of the lexicon, the latter intended both as a study of the patterns used for coining new linguistic signs, and as the quest for general patterns of polysemy of linguistic signs across languages (irrespective of the internal structure of the signs).

We will first briefly introduce issues and results of the cross-linguistic study of word-formation (3.1.1), then we will consider the issues connected with cross-linguistic comparability and the connection of the cross-linguistic study of word-formation with lexical typology (3.1.2). This will lead us to review some crucial concepts connected with the cross-linguistic study of the lexicon and the interplay of lexicon and grammar (3.1.3-3.1.5).

3.1.1 The cross-linguistic study of word-formation: issues and results

Language typology is the domain of language studies that systematically considers the structural similarities of languages irrespective of their genetic relatedness. Today, this approach to language variation can rely on a long tradition of studies and theoretical thinking spanning over two centuries, ever since the first pioneering observations made by German scholars from the turn of the nineteenth century, such as Johann Gottfried Herder, Friedrich and August von Schlegel, Wilhelm von Humboldt (Croft 2003: 45; Velupillai 2012: 1).[58] Interestingly, the first proposals for a structural classification of languages concerned the domain of

[58] For a survey of the early contributions in linguistic typology see Ramat (2010).

morphology. However, until recently, the interest in morphology within linguistic typology has mostly been directed towards inflectional exponence, or at best towards affixation in general, lumping together inflectional and derivational affixes. This becomes evident when browsing the World Atlas of Language Structures (Dryer and Haspelmath 2013), where the 12 features classified as pertaining to morphology essentially address inflection (mainly the marking of case, possessive relations, and Tense-Aspect-Mood, number and person categories on the verb),[59] while the more formally driven chapters about prefixing vs. suffixing preference (Dryer 2013) and reduplication (Rubino 2013) are no exception: the first is again explicitly devoted exclusively to inflectional morphology («This map shows the overall extent to which languages use prefixes versus suffixes in their inflectional morphology.» Dryer 2013), while in the second the problem of derivation is at best intimated in the rough outline of the range of reduplication functions («Reduplication is also used derivationally to alter word class [ ]». Rubino 2013).

In recent years, a growing interest in word-formation has enriched the cross-linguistic empirical coverage of this domain of morphology, with the publication of several overarching works that are devoted to word-formation in general (Müller *et al.* 2015a, 2015b, 2015c, 2016a, 2016b), or focus either on compounding (Scalise and Vogel 2010, Lieber and Štekauer 2009, the special issue of *Lingue e Linguaggio* presented by Bisetto 2009), or derivational morphology (Lieber and Štekauer 2014b). In the cited works, a remarkable portion of the contributions is devoted to an outline of word-formation, derivation or compounding in individual languages, or language families and branches. There are still few studies devoted to cross-linguistic surveys, both of specific word-formation phenomena and of the whole domain, with the prominent exception of Štekauer *et al.* (2012).

Lieber and Štekauer (2014a) offer an overview of the results in the research carried out on universals in derivational morphology, and examine the issues that limit its success. They note that the research on universals in word-formation (they actually focus on derivation only) has so far produced only a few results, to the point that

[59] The focus on inflection dominates in WALS in the feature areas of "Nominal categories" and "Verbal categories" as well, while the area "Lexicon" is dominated by chapters devoted to the different ways in which different languages conceptualize distinctions in the lexicon or organize pronominal systems according to different phonological patterns.

> «[o]ut of the 2,029 universals included in the [Universal Archives (Plank and Filimonova 2000)] only fifty-one are directly categorized as universals pertaining to word-formation [...]. Only five seem to us to bear directly on derivation. To these we can add a few more proposed universals discovered by searching directly for "derivational morphology" in the database, but in the end, only nine of all the assembled universals can be interpreted (somewhat generously) as focusing on synchronic word formation either in isolation or in comparison to inflection.» (Lieber and Štekauer 2014a: 781)

3.1.2 Cross-linguistic comparability and lexical typology

This study constitutes an attempt at applying a cross-linguistic perspective to a phenomenon in the domain of word-formation. This aim does not come without preliminary controversial aspects, such as obtaining cross-linguistic comparability while limiting the domain of interest of the study to a structurally identified realm such as word-formation.

Croft (2003) has defined the standard research strategy for typological research as follows:

i. Determine the particular semantic(-pragmatic) structure or situation type that one is interested in studying.
ii. Examine the morphosyntactic construction(s) or **strategies** used to **encode** that situation type.
iii. Search for dependencies between the construction(s) used for that situation and other linguistic factors: other structural features, other external functions expressed by the construction in question, or both. (Croft 2003: 14; emphasis by the author)

In the context of this study, the particular semantic-pragmatic structure is contrastive coreference – see Chapter 1, section 1.3 and Chapter 2, section 2.1 for more details – while the (word-formation) strategies used to encode it are $_{CC}$markers. We do not exclude that contrastive coreference – especially considering its strong affinity with intensifiers and reflexivity – can also be expressed by syntactic means, however, we consider contrastive coreference as a word-formation-specific meaning, based on the assumption that there is something special

about word-formation and the meanings it conveys, such as a higher degree of conventionalization and lower level of compositionality – see Chapter 2, section 2.2.8 – compared to syntactic constructions.

Another assumption that is worth putting forward concerns the nature of word-formation. In this study, word-formation is intended as a component of the lexicon, and specifically, as the sum of formal regularities that can be generalized across series of words and between different series of words defined both structurally and semantically. In this sense, the products of word-formation should be distinguished from the products of syntactic patterns – see Chapter 2, section 2.2.8. On the other hand, we do not maintain a strict subdivision of lexicon and syntax, in the sense that we assume that the lexicon can interact with syntax in several ways, e.g. by storing its products which may then participate in the general patterns of regularity of word-formation.

The details of this approach to word-formation – i.e. Construction Morphology – have been presented in Chapter 2, section 2.3. This hint at the nature of word-formation is instrumental here for linking the approach adopted in this book to the wider domain of the study of lexical typology.

Lexical typology has been defined as the «cross-linguistic and typological branch of lexicology» (Koptjevskaja-Tamm *et al.* 2007: 159), and among its main issues – especially in the narrower domain of lexical semantic typology – the following have been identified:

i. the study of «the architecture of lexical fields or semantic domains (e.g. basic vs. derived words)» (*ibid.* 161);
ii. the study of «crosslinguistically recurrent patterns of motivation, i.e., in the relations among words and lexical items in the lexicon», for example comparing «whole CLASSES or GROUPS OF WORDS where one of the classes contains words derived from, or formed on, words in the other one, and ask about the semantic relations associated with a particular word formation device (derivational patterns, compounding). The focus here is on the REGULARITIES in lexical motivation seen as an interaction of formal (morphological) and semantic aspects of motivation (cf. Koch & Marzo 2007)» (*ibid.* 162; small capitals in the original text);

It is evident how the aforementioned issues converge with the perspective on word-formation we have hinted at above.

3.1.3 Lexicalization

On a more general level, issues relevant to lexical typology concern lexicalization and lexicalization patterns, i.e. the cross-linguistic study in which meanings are expressed by a single word. These issues imply a synchronic approach to the concept of lexicalization which «refers to the extent to which there are links between conceptual representation and syntax, and how the nature of such links may be formalized» (Brinton and Traugott 2005: 18). Brinton and Traugott (2005: 19) point out that current research in this domain is concerned with gathering evidence of conflation patterns of complex conceptual structures into single lexical forms, and recognize Talmy's (1985, 2000) work on the lexicalization of motion event types as a very influential line of research.[60]

The synchronic sense of lexicalization contrasts with a diachronic one, which can be intended as the adoption of simple or complex words or of new senses into the lexicon (cfr. Blank 2001: 1603). This process of inclusion in the lexicon can correspond to the simple storage of a formation so that it can be more easily recalled later, or it can be interpreted (as is often the case) as an increase in fusion by which a regular construction becomes opaquer and more idiomatic, and the morpheme boundaries progressively coalesce so that the meaning of the word can no longer be accessed analytically, but only holistically. This view of lexicalization is advocated, for example, by Lehmann (2002: 16), who interprets it as a process which shifts items from the grammatical component (analytic, transparent, regular) to the lexical component (holistic, opaque, irregular).

In this work, we will privilege a view of lexicalization as corresponding – in line with the perspective of Construction Morphology[61] – to the ability of the lexicon to store regular or idiosyncratic constructions, words or phrases. This is in line with a view of the lexicon as a rich storage, and with the idea of a grammar-lexicon continuum – see also Chapter 2, section 2.2.8, about the boundary between compounding and phrases.

[60] Within the recent generative framework of Minimalism (Chomsky 1995, 2005), an instance of lexicalization in the synchronic sense, is also one of the stages by which the "derivation" of a linguistic expression is accomplished. In this context, lexicalization is «basically the process of turning concepts or linguistic features related to an utterance into words» (Pustejovsky and Batiukova 2019: 61).

[61] This view of the lexicon as a rich and structured inventory of both regular and idiosyncratic pairings of form and meaning (i.e. constructions) is linked to the usage-based orientation of Construction Morphology, see Masini and Audring (2019: 380-381).

In the synchronic sense, Hohenhaus (2005: 356) points out that lexicalization can be intended as the process of listing – i.e. the entering of a word into the lexicon – or the state of listedness – i.e. the property of a word of having a lexical entry in the language. He also points out, in the diachronic sense, that many studies on lexicalization stress the demotivation, i.e. the lack of compositionality which is an effect of the listedness, but it does not necessarily overlap with it, so he advocates for a distinction between idiomatization (i.e. loss of compositionality) and lexicalization (i.e. inclusion in the lexicon).

3.1.4 Grammaticalization

In diachronic sense, lexicalization – especially if intended as an increase in fusion à la Lehmann – is often put in contrast with grammaticalization, i.e. a somewhat opposite phenomenon by which items originally included in the lexicon progressively lose lexical meaning and independence.

Lehmann (2002) argues that folk etymology, i.e. the process «[b]estowing structure onto a hitherto opaque expression» (Lehmann 2002: 16), is actually the opposite of lexicalization, while the opposite of grammaticalization would be degrammaticalization, i.e. the process of «[g]iving autonomy to a hitherto dependent expression» (Lehmann 2002: 17). Note, however, that lexicalization and degrammaticalization share several aspects, even if they do not completely overlap (s. discussion in Brinton 2012: 1588-1590).

The development of a word-formation pattern, especially the development of derivational affixes out of syntactically free words, has been considered an example of grammaticalization. Hopper and Traugott (2003: 40-41) consider the development of the English suffix *-hood* from Old English *had* 'person, condition rank' as a case of grammaticalization, given the fact that a new grammatical meaning is added to the grammar and *-hood* simply forms abstract nouns from nouns. They contrast that example with that of e.g. *bosun/boatswain* lit. 'boat lad' – which they define as a case of lexicalization – where the compound simply loses its compositionality, while the second element, differently from *had*, has not been reanalyzed as a grammatical morpheme (Hopper and Traugott 2003: 58).

This stance is not, however, always shared among scholars. For example, the same example of English *-hood* has been interpreted as a case of lexicalization in Blank (2001), who, however, concedes

> «If, however, one prefers to see word-formation as the "grammaticalization of the lexicon", the change of an autonomous word into an affix (and from free compounding into a rule-based process) is rather a kind of grammaticalization.» (Blank 2001: 1603)

We have seen already in Chapter 2, section 2.2.1-2.2.2, that derivation is somewhere halfway between inflection and derivation as far as the meaning of derivational operations is concerned. If compared to the meanings of inflectional operations, they appear as lexical; if compared to the meanings of compound units, they appear as grammatical. This suggests a cline which we can consider as one of the aspects of the lexicon-grammar continuum. On the one hand, we have an increase in lexical content, moving from inflectional exponence to derivational affixes, and from there to compound units and free words. On the other hand, from the perspective of Construction Grammar, we have considered syntactic constructions as closer to grammar, while an increasing integration into the lexicon (in the sense of listedness, but also of idiomaticity) is obtained by moving from syntactic constructions towards compounds and derivational patterns. Within word-formation and from the perspective of Construction Morphology, we have observed that the hierarchical lexicon allows for the specification of idiosyncratic properties at any level of the hierarchy. This also generates a continuum between higher, more abstract, i.e. more grammatical, nodes, and lower, more specified, i.e. more lexical nodes.

In this picture, inflection appears somewhat misplaced because it operates on a rather grammatical level, in the sense that it can be viewed as interacting more with syntactic constructions compared to word-formation, but its products, i.e. grammatical words, are in many cases considered as stored in the lexicon, especially in the case of languages that belong to the fusional type.

As the object of this study, $_{CC}$markers have a rather interesting nature as far as their position on a grammaticalization cline is concerned. From the semantic point of view, both Italian *auto-* and German *selbst-* are characterized by a rather grammatical contribution. However, from a formal point of view, German *selbst-* is far less grammaticalized than *auto-*, given the possibility to conceive of it as an instance of a NN compound.

Hüning and Booij (2014) highlight how applying the concept of grammaticalization to the domain of word-formation can be counterproductive, and

> «[i]t is especially the dichotomy of "the lexical" vs. "the grammatical" that turns out to be inadequate for a proper account of word formation phenomena, since in word formation we have always to deal with both aspects.» (Hüning and Booij 2014: 600)

They draw on Trousdale and Norde's (2013: 36) concept of constructionalization, defined «as a sequence of changes in the form and meaning poles of a construction, whereby new formal configurations come to serve particular functions, and to encode new meanings». The advantage of referring to constructionalization, instead of contrasting it with grammaticalization or lexicalization, corresponds to the fact that the constructional approach «suggests not a cline, but a taxonomic network of related constructions» (Trousdale 2008: 172), as the ones that we have also seen in the case of *auto-* and *selbst-* in Chapter 2, section 2.3.4. In this vein, Hüning and Booij conclude that

> «[t]he development of derivational affixes from compound constituents is primarily a case of constructionalization, the rise of a morphological construction, and morphological change can adequately be analyzed as constructional change at the word level». (Hüning and Booij 2014: 600)

3.1.5 Areal tendencies

The concept of grammaticalization, be it relevant for word-formation or not, has been applied in recent years not only to diachronic changes affecting the grammar of a specific language as an internal development, but also to changes that occur in one language under the pressure of patterns found in another language which the first one comes in contact with. Contact-induced grammaticalization is tightly linked with areal phenomena and to the concept of linguistic area, or *Sprachbund*. It has been recognized that European languages, because of the shared history of long-standing cultural, social and political influences, have ended up sharing linguistic features that emphasize their similarity and, conversely, contrast them to the behaviour of the languages spoken in the rest of the world. This is the concept of Standard Average European, which has guided the EuroTyp project and yielded a series of contributions with the aim of providing a typological profile of Europe and the European linguistic landscape.

The borrowing of single lexical items (be they lexical or functional) is the most common type of influence in language contact (MAT or matter borrowing), often implying a lower level of knowledge of the source language by speakers of the recipient language. However, when contact increases, it is not rare that a larger stock of borrowings also leads to the adoption of morphological patterns from a source language (PAT or pattern borrowing).[62]

In the context of the study of $_{CC}$markers, contact surely plays a prominent role – see Chapter 2, section 2.2.4-5 above, and the case of Italian *auto-* and its origin in the neoclassical lexical stock – whose influence on the form and meaning of European $_{CC}$markers will be evaluated in Chapter 6.

3.2 Sample and methodology

This study is a first cross-linguistic survey in the domain of contrastive coreference. It aims at exploring this area in order to give a first account of the phenomenon that can be used as the basis for further future research. The language sample used in this study (see Table 3) is in no way representative of the languages of the world: it is genetically biased towards Indo-European languages, areally biased towards Europe, and typologically biased towards SOV and prepositional languages belonging in most of the cases to the morphological type of synthetic-fusional languages.

The sample can be viewed as a convenience sample: European and especially Indo-European languages are among the most documented languages in the world, and this documentation also covers the description of their word-formation devices. On the other hand, the Eurocentricity of the sample is justified by the fact that the aim of the study includes a synchronic as well as a diachronic account: the availability of extensive documentation covers several centuries (for most languages in the sample), if not millennia (for the Italic/Romance, German, the Slavic branch and Greek). In this sense, the biases of the sample are a prerequisite for undertaking a diachronic exploration of the data that will be attempted in Chapter 6.

Beyond the issue of availability of documentation, the focus on Europe is also

[62] See Gardani (2020) for a survey of the problem in the domain of morphology.

rooted in the fact that this continent has been recognized as a linguistic area characterized by numerous structural affinities displayed by European languages. Haspelmath (1998, 2001) surveys the features of Standard Average European (SAE) in an attempt to establish when the attested grammatical commonalities might have spread across Europe.

He proposes that the transition between Antiquity and the Middle Ages, a time of great migrations across Europe, might be the most probable time frame at which a number of features of SAE established themselves – at least articles, the 'have'-perfect, the participial passive, anticausatives, negative indefinites, nominative experiencers and verb fronting, i.e. seven out of the twelve features considered in Haspelmath (2001).

He also observes that:

> «A time depth of 300–500 years is not sufficient to account for grammatical commonalities of the kind to be discussed below. If lexical similarities between the European languages are discussed–for instance neoclassical compounding (*socio-*/*paleo-*/*ortho-*/*demo-*, *-graphy*/*-logy*/*-cracy*, etc.) or idiomatic structure (e.g. *ivory tower*/*torre d'avorio*/*Elfenbeinturm*, as poor as a *church mouse*/*pauvre comme un rat d'église*/*arm wie eine Kirchenmaus*)—then the last several centuries are the appropriate time frame for explaining the historical links […].» (Haspelmath 1998: 272)

3.2.1 Genetic profile of the languages

The language sample used in this study mostly overlaps with the sample used in Haspelmath (1998, 2001). The languages included in the two samples are listed in Table 2.1, along with the abbreviations used to refer to each of them, and their genealogical classification according to *Glottolog 4.5*.[63]

Most languages belong to the Indo-European language family, and in most cases the list exhausts the major languages of each branch – i.e. those which are

[63] *Glottolog 4.5* (Hammarström *et al.* 2021) is a website created at the initiative of the Max Planck Institute for Evolutionary Anthropology of Leipzig, and contains information about different languages and language families of the world, especially genealogical classification and bibliographical information accompanied by a comprehensive tagging of languages and dialects through a unique and stable identifier (called Glottocode). *Glottolog 4.5* is actively maintained and regularly updated by its creators and with the help of users (especially expert linguists).

official languages of a European state. Obviously, not all languages in the sample share the same degree of vitality. For example, only a few of the languages in the sample have more than 100 million speakers, and some of them have most of their speakers outside Europe (e.g. English, Spanish, Portuguese). At the other extreme, some languages are minority languages or even endangered languages, for example, all languages representing the Celtic branch of the Indo-European language family are minority languages, spoken alongside another Indo-European language (English or French). This difference concerning vitality influences the availability of sources as well as their quality and richness: impressionistically, we can expect that the smaller a language is, the less vital it will be, and, consequently, it will have a smaller lexicon and less active word-formation.

We will leave this aspect of the correlation between vitality and lexical and morphological richness as a suggestion for future research. It is, however, important to highlight this as one of the issues that might affect the results of this study.

Table 3. The language sample: genetic profile and comparison with the sample in Haspelmath (1998, 2001)

Language	Abbreviation	Language Family	Classification below top-level (adapted from *Glottolog 4.5)*	Haspelmath (1998, 2001)	This book
Italian	It	Indo-European	Italic, Romance	x	x
North Italian	NIt	Indo-European	Italic, Romance	x	
Sardinian	Srd	Indo-European	Italic, Romance	x	
French	Fr	Indo-European	Italic, Romance	x	x
Spanish	Spn	Indo-European	Italic, Romance	x	x
Portuguese	Prt	Indo-European	Italic, Romance	x	x
Romanian	Rom	Indo-European	Italic, Romance	x	x
German	Grm	Indo-European	Germanic, West G.	x	x
English	Eng	Indo-European	Germanic, West G.	x	x
Dutch	Dut	Indo-European	Germanic, West G.	x	x
Norwegian	Nor	Indo-European	Germanic, North G.	x	
Danish	Dan	Indo-European	Germanic, North G.		x
Swedish	Swd	Indo-European	Germanic, North G.	x	x
Icelandic	Ice	Indo-European	Germanic, North G.	x	x

Language	Abbreviation	Language Family	Classification below top-level (adapted from *Glottolog 4.5)*	Haspelmath (1998, 2001)	This book
Irish Gaelic	Ir	Indo-European	Celtic, Insular, Goidelic	x	x
Scottish Gaelic	Sc	Indo-European	Celtic, Insular, Goidelic		x
Welsh	Wel	Indo-European	Celtic, Insular, Brythonic	x	x
Breton	Brt	Indo-European	Celtic, Insular, Brythonic	x	x
Bulgarian	Blg	Indo-European	Balto-Slavic, Slavic, Eastern South S.	x	
Serbian and Croatian	SCr	Indo-European	Balto-Slavic, Slavic, Western South S.	x	
Serbian	Srb	Indo-European	Balto-Slavic, Slavic, Western South S.		x
Croatian	Cro	Indo-European	Balto-Slavic, Slavic, Western South S.		x
Slovene	Sln	Indo-European	Balto-Slavic, Slavic, Western South S.	x	x
Czech	Cz	Indo-European	Balto-Slavic, Slavic, West S.	x	x
Polish	Pol	Indo-European	Balto-Slavic, Slavic, West S.	x	x
Russian	Rus	Indo-European	Balto-Slavic, Slavic, East S.	x	x
Ukrainian	Ukr	Indo-European	Balto-Slavic, Slavic, East S.	x	
Lithuanian	Lit	Indo-European	Balto-Slavic, Eastern Baltic	x	x
Latvian	Ltv	Indo-European	Balto-Slavic, Eastern Baltic	x	x
Albanian	Alb	Indo-European	–	x	x
Modern Greek	Grk	Indo-European	–	x	x
Armenian	Arm	Indo-European	–	x	
Finnish	Fin	Uralic	Finnic	x	x

Language	Abbreviation	Language Family	Classification below top-level (adapted from *Glottolog 4.5)*	Haspelmath (1998, 2001)	This book
Estonian	Est	Uralic	Finnic	x	x
Hungarian	Hng	Uralic	Ugric	x	x
Komi	Kom	Uralic	Permian	x	
Udmurt	Udm	Uralic	Permian	x	
Mari	Mar	Uralic	–		
Nenets	Nnts	Uralic	Samoyedic	x	
Turkish	Trk	Turkic	Oghuz, West O.	x	x
Uzbek	Uzb	Turkic	Turkestan Turkic		x
Tatar	Tat	Turkic	Kipchak, North K.	x	
Georgian	Grg	Kartvelian	Georgian-Zan	x	x
Lezgian	Lzg	Nakh-Daghes-tanian	Daghestanian, Lezgic, Samur	x	
Basque	Bsq	–	–	x	x
Maltese	Mlt	Afro-Asiatic	Semitic, Arabian	x	x

In Figure 8, the languages of Haspelmath's (1998) sample are plotted so that their distribution in the figure corresponds to the geographical location of these languages relative to one another. Figure 8 also subdivides the languages according to their degree of appurtenance to Standard Average European, with nucleus languages sharing most of the relevant features, and core and periphery languages sharing a decreasing number of features. Other languages have been considered by Haspelmath as external to Standard Average European, and interestingly, these include a branch of Indo-European languages, i.e. Celtic languages.

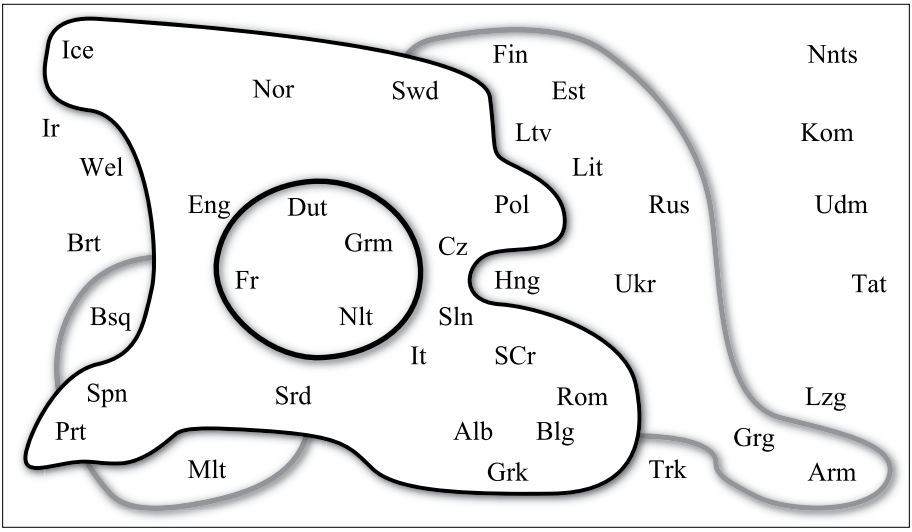


Figure 8. Standard Average European (nucleus, core, periphery; Haspelmath 1998: 273)

On the basis of nine of the twelve features considered in Haspelmath (1998), Haspelmath (2001) defines a slightly different picture, according to which only French and German constitute the nucleus of Standard Average European, while the other languages of the sample display a decreasing number of features, the ones with two or less features being considered as external to the linguistic area – see Figure 9.

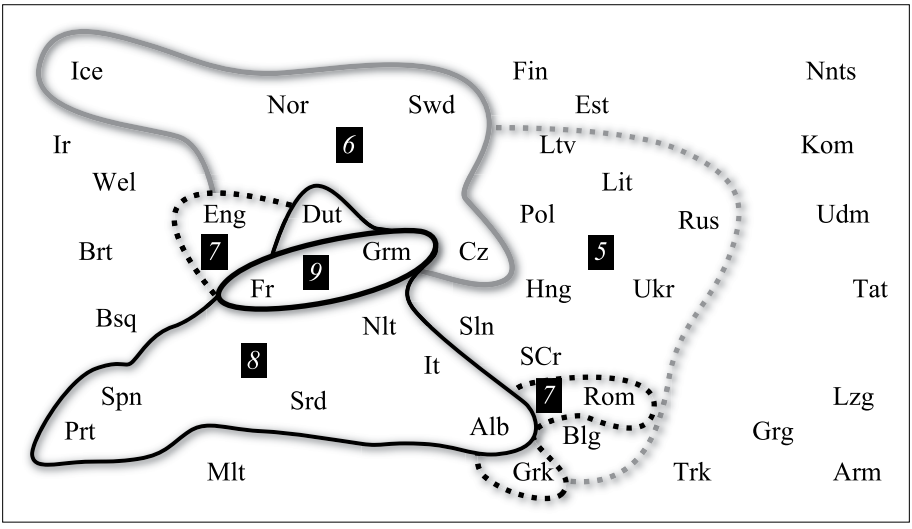


Figure 9. Standard Average European (decreasing number of shared features; Haspelmath 2001)

Among the features considered by Haspelmath (1998, 2001), one is of particular relevance for our study of $_{CC}$markers: the differentiation between intnsifiers and reflexive markers in European languages – see Figure 10.

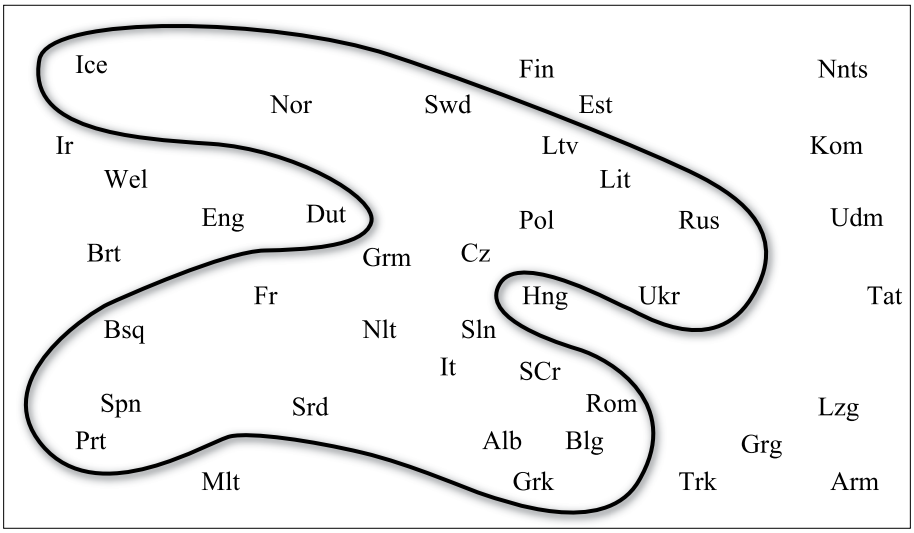


Figure 10. Intensifier-reflexive differentiation (Haspelmath 2001: 1501)

In the next chapter, we will survey the sources of $_{CC}$markers in around 30 languages of Europe, and we will discuss the differentiation of intensifiers and reflexives in more detail – see Chapter 4, section 4.1.: as we have already observed in the previous chapter about German, intensifiers can be one of the possible sources of $_{CC}$markers.

3.2.2 Typological profile of the languages

In Table 4, we have reported some of the typological features which characterize the languages included in our sample.[64]

The features considered correspond to the five chapters of the WALS (Dryer and Haspelmath 2013): Order of Subject, Object and Verb (WALS 81, Dryer 2013b), Order of Adposition and Noun Phrase (WALS 85, Dryer 2013c), Order of Gen-

[64] I have actually reported the values for all the languages included in Table 1, i.e. all the languages of Haspelmath's (1998, 2001) sample.

itive and Noun (WALS 86, Dryer 2013d), Order of Adjective and Noun (WALS 87, Dryer 2013e), Prefixing vs. Suffixing in Inflectional Morphology (WALS 26, Dryer 2013a). I considered these features because they might be good predictors of the relative position of head and modifier in compounds, but they might also correlate with the presence of prefixes in a language.[65]

Table 4. Typological classification of the languages of the sample[66]

Language	WALS 81	WALS 85	WALS 86	WALS 87	WALS 26
It	(S)VO	Prep	N-Gen	N-Adj	Strongly suff.
NIt	n.a.	n.a.	n.a.	n.a.	n.a.
Srd	(S)VO	n.a.	n.a.	N-Adj	Strongly suff.
Fr	(S)VO	Prep	N-Gen	N-Adj	Strongly suff.
Spn	(S)VO	Prep	N-Gen	N-Adj	Strongly suff.
Prt	(S)VO	Prep	N-Gen	N-Adj	Strongly suff.
Rom	(S)VO	Prep	N-Gen	N-Adj	Strongly suff.
Grm	(S)VO/(S)OV	Prep	N-Gen	Adj-N	Strongly suff.
Eng	(S)VO	Prep	N.D.O.	Adj-N	Strongly suff.
Dut	(S)VO/(S)OV	Prep	N-Gen	Adj-N	Strongly suff.
Nor	(S)VO	Prep	N.D.O.	Adj-N	Strongly suff.
Dan	(S)VO	Prep	Gen-N	Adj-N	Strongly suff.
Swd	(S)VO	Prep	Gen-N	Adj-N	Strongly suff.
Ice	(S)VO	Prep	N-Gen	Adj-N	Strongly suff.
Ir	V(S)O	Prep	N-Gen	N-Adj	Equal pref. and suff.
Sc	V(S)O	Prep	N-Gen	N-Adj	Weakly suff.
Wel	V(S)O	Prep	N-Gen	N-Adj	Strongly suff.
Brt	(S)VO	Prep	N-Gen	N-Adj	Weakly suff.
Blg	(S)VO	Prep	N.D.O.	Adj-N	Strongly suff.
SCr	(S)VO	Prep	N.D.O.	Adj-N	Strongly suff.
Sln	(S)VO	Prep	N.D.O.	Adj-N	Strongly suff.

[65] See Hawkins and Gilligan (1988) about the well-known issue of the suffixing preference in the languages of the world. It should be noted that in their study they consider mostly inflectional affixes and, more generally, they do not distinguish between inflectional and derivational affixes.

[66] Abbreviations: Adj = adjective; Gen = genitive; N = noun; n.a. = information not available through WALS; N.D.O. = no dominant order; O = object; pref. = prefixing; Prep = preposition; Postp = postposition; S = subject; suff. = suffixing; V = verb

Language	WALS 81	WALS 85	WALS 86	WALS 87	WALS 26
Cz	(S)VO	Prep	N.D.O.	Adj-N	Weakly suff.
Pol	(S)VO	Prep	N-Gen	Adj-N	Strongly suff.
Rus	(S)VO	Prep	N-Gen	Adj-N	Strongly suff.
Ukr	(S)VO	Prep	n.a.	Adj-N	Strongly suff.
Lit	(S)VO	Prep	Gen-N	Adj-N	Strongly suff.
Ltv	(S)VO	Prep	Gen-N	Adj-N	Weakly suff.
Alb	(S)VO	Prep	N-Gen	N-Adj	Strongly suff.
Grk	(S)VO/V(S)O	Prep	N-Gen	Adj-N	Strongly suff.
Arm	(S)VO/(S)OV*	Postp	Gen-N	Adj-N	Strongly suff.
Fin	(S)VO	Postp	Gen-N	Adj-N	Strongly suff.
Est	(S)VO	Postp	Gen-N	Adj-N	Strongly suff.
Hng	(S)VO/(S)OV	Postp	Gen-N	Adj-N	Strongly suff.
Kom	(S)VO	Postp	Gen-N	Adj-N	Strongly suff.
Udm	(S)OV	Postp	Gen-N	Adj-N	Strongly suff.
Mar	n.a.	Postp	Gen-N	Adj-N	Strongly suff.
Nnts	(S)OV	Postp	Gen-N	Adj-N	Strongly suff.
Trk	(S)OV	Postp	Gen-N	Adj-N	Strongly suff.
Uzb	(S)OV	n.a.	Gen-N	Adj-N	Strongly suff.
Tat	(S)OV	Postp	Gen-N	Adj-N	Strongly suff.
Grg	(S)OV	Postp	Gen-N	Adj-N	Weakly suff.
Lzg	(S)OV	Postp	Gen-N	Adj-N	Strongly suff.
Bsq	(S)OV	Postp	Gen-N	N-Adj	Equal pref. and suff.
Mlt	(S)VO	Prep	N-Gen	N-Adj	n.a.

It is clear from Table 4 that the languages of this sample are also typologically rather homogeneous, since most of them have the SOV word order and prepositions, and are strongly suffixing. The situation is a bit more varied as far as the relative order of Genitive and Noun and of Adjective and Noun are concerned, but the value Adjective-Noun evidently dominates.

The values in Table 4 (with the exception of the affixing preference) can be interpreted as binary features, and can consequently be used to create a graph such as the one in Figure 11, which has been generated by using the SplitsTree program (Huson and Bryant 2006).

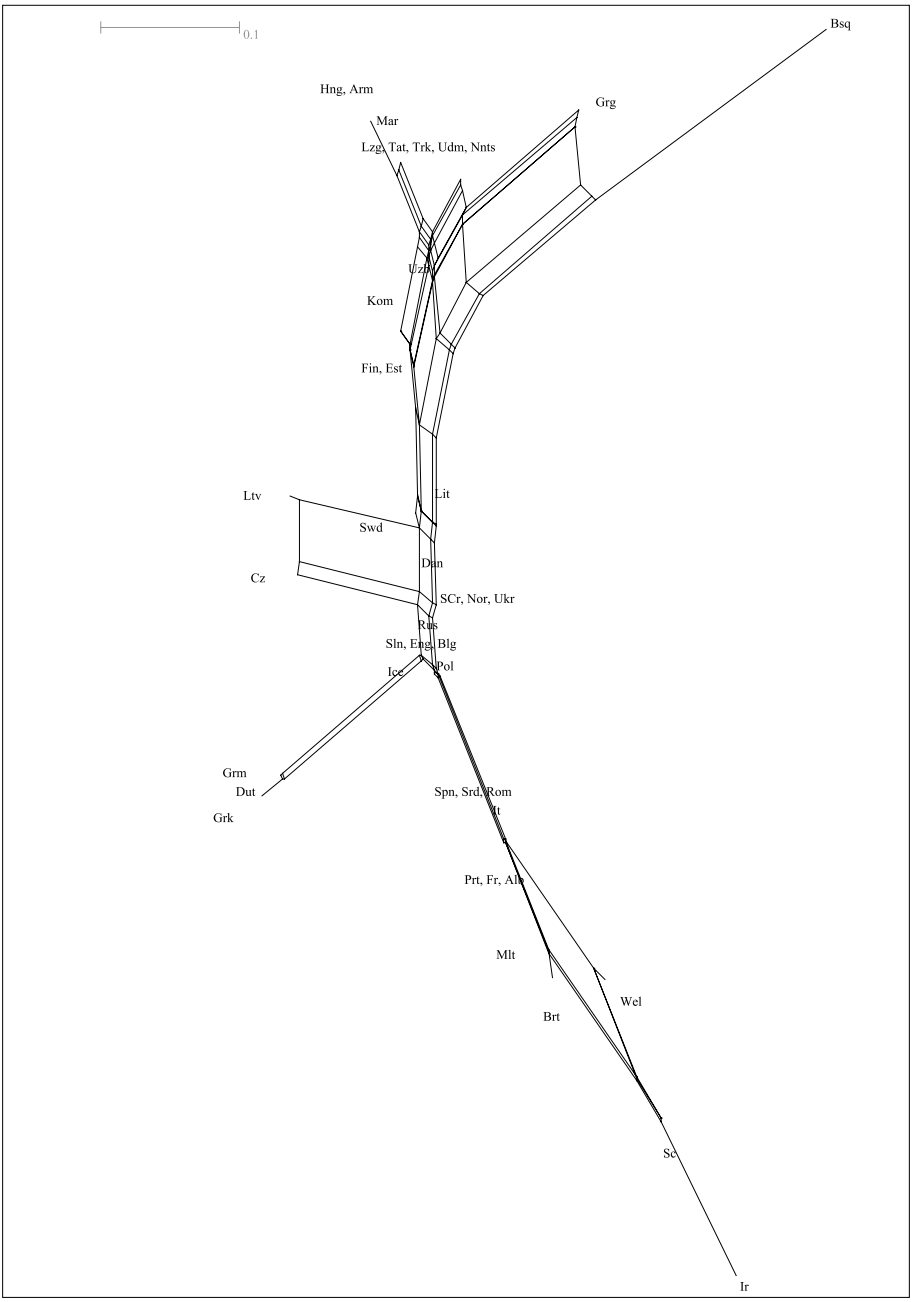


Figure 11. SplitsTree plot of distances between the languages of the sample based on the word order features of Table 4

The graph plots the Hamming distance between pairs of languages, based on the difference in the presence vs. absence of each of the features from Table 4.

Romance languages are in the branch of the graph that is pointing downwards, along with Celtic languages (Irish constitutes the extreme low-end point of the graph), Albanian and Maltese. Germanic languages share the central part of the graph with Slavic and Baltic languages. Greek, German and Dutch constitute the extreme part of the central left branch of the graph. Uralic languages are at the top of the graph, along with Turkic languages, Armenian and Georgian. Basque occupies the upper right extreme of the graph.

3.2.3 Morphological profile of the languages: prefixation in WF and compounding

The last profile included in this chapter concerns the morphological profile of the languages of the sample, focusing on word-formation patterns, i.e. the presence of prefixation in word-formation, its input categories, and the position of the head in compounds. This profile is relevant, because it represents the highest level of abstraction in the hierarchical lexicon of these languages as far as prefixation and compounding are concerned. For example, knowing that Italian, like Romance languages in general, has a preference for left-headed compounds, makes right-headed compounds less likely to be a productive means of word-formation in Romance languages, or possibly a more constrained device. Conversely, knowing that most Uralic languages display a marginal use of prefixation, makes its presence as a candidate for prefixation more remarkable, since in the hierarchical lexicon prefixation is likely to be very constrained and unproductive.

Table 5. Prefixation in WF and position of the head in compounding (based on the sketches in Müller *et al.* 2016a, 2016b)

Language	Prefixes in derivational morphology	Input categories of prefixes	Prevalent position of the head in composition	Source
It	yes	N, A, V	left (marginally right)	Rainer (2016b)
NIt	-	-	-	-
Srd	marginal	-	(left)	Pinto (2016)
Fr	yes	N, A, V (?)	left	Floricic (2016)

Language	Prefixes in derivational morphology	Input categories of prefixes	Prevalent position of the head in composition	Source
Spn	yes	N, A, V	left (marginally right)	Rainer (2016a)
Prt	yes	N, A, V	left (marginally right)	Pöll (2016)
Rom	yes	N, A, V	left (marginally right)	Grossmann (2016)
Grm	yes	N, A, especially V	right	Barz (2016)
Eng	yes	N, A, V	right	Plag (2016)
Dut	yes	N, A, especially V	right	Booij (2016)
Nor	yes	N, A, especially V (less prominent in Nynorsk)	right	Askedal (2016)
Dan	yes	N, A, especially V	right	Götsche (2016)
Swd	yes	N, A, especially V (mostly Low German loans)	right	Kotcheva (2016)
Ice	yes	N, A, V	right	Indriðason (2016)
Ir	yes (?)	(input V, output A)	right	Ó Curnáin (2016)
Sc	-	-	-	-
Wel	yes	N, A, V	right	Russell (2016)
Brt	yes	N, A, V	right (older), left (newer)	Ternes (2016)
Blg	yes	N (loans), A, especially V	right	Avramova and Baltova (2016)
Srb	yes	N, A, especially V	right	Ćorić (2016)
Cro	yes	N, A, especially V	right	Grčević (2016)
Sln	yes	N, A, especially V	right	Stramljič Breznik (2016)

Language	Prefixes in derivational morphology	Input categories of prefixes	Prevalent position of the head in composition	Source
Cz	yes	N, A, especially V	right	Bozděchová (2016)
Pol	yes	N (loans), A, especially V	right (marginally left)	Nagórko (2016)
Rus	yes	N, A, especially V	right	Uluhanov (2016)
Ukr	yes	N, A, especially V	right	Karpilovska (2016)
Lit	yes	N and A (marginal), especially V	right	Studžia (2016)
Ltv	yes	N (marginal), A, especially V	right	Navickaitė-Klišauskienė (2016)
Alb	yes	N, A, especially V	left (older), right (newer)	Genesin and Matzinger (2016)
Grk	yes	N, A, V	right	Ralli (2016)
Arm	-	-	-	-
Fin	marginal, bordering compounding	-	right	Pitkänen-Heikkilä (2016)
Est	marginal, bordering compounding	-	right	Kerge (2016)
Hng	yes	exclusively particle Vs	right	Kiefer (2016)
Kom	marginal	-	right	Ferjes (2016)
Udm	marginal	-	right	
Mar	marginal	-	right	Riese (2016)
Nnts	marginal	-	right	Wagner-Nagy (2016)
Trk	marginal	-	right	Wilkens (2016)
Uzb	-	-	-	-
Tat	marginal	-	right	Károly (2016)

Language	Prefixes in derivational morphology	Input categories of prefixes	Prevalent position of the head in composition	Source
Grg	-	-	-	-
Lzg	-	-	-	-
Bsq	marginal	-	right	Artiagoitia *et al.* (2016)
Mlt	marginal (in loan-words)	-	left (construct state)	Brincat and Mifsud (2016)

Most languages of the sample display prefixation, even if some languages appear to have a richer stock of preverbs combining with verbal bases (e.g. Slavic languages, Baltic languages, most Germanic languages, Hungarian). Other Uralic languages, Turkic languages (Turkish and Tatar) and Maltese appear as having only a marginal use of prefixes in word-formation. It must be noted that Maltese is the only language of the sample that has an introflexive, and consequently, a non-concatenative morphology. However, the case of Maltese is complexified by contact, in that a significant portion of the lexicon is constituted by Romance loanwords displaying a concatenative morphology instead.[67]

Right-headed compounds constitute the most widespread type of compounds in the sample, also expanding (most likely because of contact) to languages preferring left-headed compounds (e.g. Romance languages and Albanian). The only example of a language moving from right-headed to left-headed compounds is Breton, probably due to the pressure of French.

3.2.4 Data: sources and issues

The main data sources for this study have been dictionaries. On the one hand, the scope of the study, i.e. $_{CC}$markers, is very restricted, corresponding not to a class of markers (affixes or compound units), but maybe to just one marker occurring in a few formations. On the other hand, the quality and depth of language descriptions is a well-known issue in cross-linguistic research. This issue, called the "bibliographic bias" by Bakker (2010), crucially affects the domain of word-formation, a domain

[67] Brincat and Mifsud (2016: 3349) illustrate this situation by reporting how the majority – around 21 thousand out of 41 thousand entries in Aquilina's (1987-1990) dictionaries – are of Sicilian and Italian origin.

that is frequently left out of focus – as Lieber and Štekauer (2014a: 778) point out; see above in 3.1.1 as well – as opposed to inflectional morphology.

The alternatives to the use of dictionaries could have included a corpus-based investigation or gathering data through elicitation. Both methodologies would have been hard to undertake, either because using a corpus implies a rather extensive knowledge of individual languages, or because data elicitation is not ideal for obtaining the amount of data (especially in terms of width of the lexical coverage) needed for investigating word-formation.

However, the use of dictionaries, which has a long tradition in the study of word-formation, is accompanied by several issues connected to the nature of lexicographic sources. Lieber (2014: 88) discusses some of the shortcomings of the use of dictionaries in the study of word-formation. She points out that dictionaries are biased toward idiosyncratic and high-frequency forms, while low-frequency, regular and compositional forms, which are most likely not finding their way into a dictionary, are on the other hand of high interest for the morphologist. Furthermore, even though a form is not included in a dictionary, it does not mean that it does not exist. However, this is often the conclusion to which a researcher might be led by acknowledging the absence of a pattern from a lexicographic source.

In addition, as we have already mentioned earlier, the languages of the sample do not have the same status and vitality, and we expect that this affects the productivity of a pattern such as that of $_{CC}$markers, which often occurs in terminologies and scientific domains.

Bearing these limitations in mind, we have gathered data from bilingual and, when possible, monolingual dictionaries, harvesting the data usually in a two-step procedure. Starting from Italian and German (or English and other languages in some cases), we gathered the translation equivalents of $_{CC}$formations with the known $_{CC}$markers of these languages in the other languages of the sample, i.e. target languages, applying a sort of onomasiological methodology. The result of this first step has been a list of candidate $_{CC}$markers for the target languages. The second step involved finding all the words in the target language that display the candidate $_{CC}$markers found in the first step. The aim of this second step, in the vein of semasiological methodology, was to provide access to a wider scope of uses of the candidate $_{CC}$markers, and ideally, produce a wider picture, less constrained by the limitations imposed by the behavior of the $_{CC}$markers in the languages constituting the point of departure.

3.3 Conclusion

In this chapter, we have described the issues of the study of word-formation cross-linguistically. We have shown that certain issues of cross-linguistic research can be overcome by implementing some of the concepts introduced in the previous chapters, such as the use of contrastive coreference as the semantic-pragmatic structure which constitutes the focus of this cross-linguistic investigation. On the other hand, we have shown how Construction Morphology can be fruitfully used as a theoretical framework, capturing aspects of concepts such as lexicalization, grammaticalization and contact that are relevant in cross-linguistic studies.

In the second part of the chapter, we have introduced the sample, and we have discussed its profile from a genetic, typological and morphological point of view. Finally, we have discussed the issues of using dictionaries in the study of word-formation, and we have exposed the methodology used for gathering the data.

In the next three chapters, we will come at the empirical core of the study with a discussion of $_{CC}$markers in the languages of the sample described in this chapter.

4

LEXICAL SOURCES OF CONTRASTIVE COREFERENCE MARKERS

In this chapter, we will consider the markers that in the different languages of the sample perform the function of contrastive coreference. The objective of the chapter is to determine as comprehensively as possible the class of word-formation processes to which $_{CC}$markers in the sample belong. In order to achieve this, we will first of all define the variety of free elements that have syntactic functions similar to those performed by $_{CC}$markers in word-formation. Our hypothesis is that the best candidates for a $_{CC}$marker are elements belonging to the intensifier-reflexive continuum, which we will introduce in 4.1. In 4.2, we will then consider each language of the sample following the genetic classification, and we will examine which elements are used to convey contrastive coreference, and whether they correspond to syntactically free elements on the intensifier-reflexive continuum. We will conclude the chapter by summarizing our survey of the lexical sources of $_{CC}$markers in the languages of Europe in 4.3, where we will show how $_{CC}$markers essentially constitute a family of non-prototypical word-formation processes that cannot be easily attributed to either derivation or compounding.

4.1 The intensifier-reflexive continuum

In Chapter 1, we have discussed the functions of Italian *auto-* and German *selbst-* and we have referred to reflexivity. While considering the functions of the free element *selbst* in Chapter 2, we have also mentioned the class of intensifiers. In this paragraph, we will show how the domain of reflexivity and that of intensifiers – characterized by specific functions, constructions and markers – are related to each other as they belong to the same grammaticalization cline. In the literature, certain specific parts of the lexicon have been recognized as a privileged source of intensifiers. The latter are in turn the reservoir from which languages draw to renew markers of reflexivity. If we exclude the less grammaticalized area of the continuum, i.e. that of the lexical sources of the intensifiers, its remaining parts are gradually characterized by an increasingly high degree of syntactic dependence as they are integrated into more or less fixed constructions. Given that the objective of this section is to identify the morphemes that can be used in complex morphological formations to convey contrastive coreference, in what follows we will not provide a series of criteria for identifying and finely distinguishing between different functions of the continuum. As a matter of fact, the

criteria needed to operate such a fine-grained distinction imply on the one hand the definition of syntactic constructions in which the relevant functions emerge. On the other hand, the aforementioned criteria also include the identification of specific forms of the paradigm of a marker. However, in word-formation, units corresponding to syntactically free elements usually appear in a form that corresponds to the root or to a specific base, which is then used in a (long) series of formations. Moreover, we will see how some forms used in word-formation can be traced back to different functions from the continuum, consequently limiting the relevance of their fine-grained identification.

In (35), we give a version of the intensifiers-reflexive continuum as presented in Heine (1999: 7) and based on the data from a study on 62 African languages.[68]

(35) NOMI > EMPH > REFL > RECI > MIDD > PASS

As can be seen, the continuum develops beyond the markers of reflexivity and reaches the territory of the middle and of the passive. For example, consider in (36) the Italian pronoun *si*, whose impersonal function can be regarded as a passivization strategy:[69]

(36) a.	*In*	*questo*	*negozio*	***si***	***vendono***	*libri*
	in	this	shop	IMPR	sell.PRS.3PL	book-PL
b.	*In*	*questo*	*negozio*	***vengono***	***venduti***	*libri*
	in	this	shop	come.PRS.3PL	sell.PST.PTCP-M.PL	book-PL

'Books are sold in this shop'

Starting from the left side of the continuum, Heine points out how the main lexical source of reflexives and reciprocals are the terms for BODY. Among other terms, the most frequent are those for HEAD and, subsequently, OWNER, COMPANION, LIFE, RELATIVE, SOUL, PERSON. From the terms indicating BODY and HEAD, markers can be derived that run along the entire continuum shown in (35). For other terms, however, restrictions emerge from the data of the study: in particular, OWNER is

[68] Labels: NOMI = nominal meaning, EMPH = emphatic reflexive, REFL = reflexive, RECI = reciprocal, MIDD = middle, PASS = passive. It must be noted that Heine (1999) deals with the polysemy between the reciprocal and the reflexive.

[69] On the grammaticalization of *si* in Italian from reflexive to impersonal see Cennamo (1993).

exclusively a source of emphatic reflexives, while COMPANION and RELATIVE only constitute the source of markers of reciprocity.

Heine observes that if a marker on the continuum has among its functions that of an emphatic pronoun/intensifier, it is unlikely to have an affixal character. If, on the other hand, a marker belongs to the class of nouns, it is unlikely that it will also be able to display functions, such as the passive, that are instead characteristic of affixes found at the right end of this grammaticalization path (Heine 1999: 18).[70]

Another study focusing on the origin of markers used in reflexive constructions in 150 languages of the world (Schladt 1999), confirms the predominance of the terms indicating parts of the body, and of the terms BODY and HEAD as sources of markers of reflexivity.[71] To this group of sources Schladt (1999) adds others, including those indicating the SELF or PERSON, anaphoric pronouns and emphatic pronouns (i.e. intensifiers). As Schladt himself notes, the distinction between SELF/PERSON and emphatic pronouns is often difficult to support, especially in the cases in which the same terms cover the functions of the former and the latter; the distinction is based solely on the availability of information provided by the authors of the grammars of individual languages (Schladt 1999: 106).

Now that we have surveyed the variety of sources on which languages draw to renew the continuum of intensifiers and reflexives, the core area of the continuum can be further examined. Heine (1999) and Schladt (1999) use the terms "emphatic pronouns" and "emphatic reflexive": in this area some clarifications are needed, firstly regarding emphatic pronouns. König and Gast (2006) attempt to systematize the domain of «expressions like English *himself/herself/itself* when they are used in an adjunct position (e.g., *the president himself*), [like] Latin *ipse/ipsa/ipsum*, Italian *stesso/stessa*, Russian *sam/samá/samó*, German *selbst*, etc., and their counterparts in other languages» (König and Gast 2006: 223). They point out

[70] In this regard, see also one of the cross-linguistic generalizations proposed in König and Gast (2006): «If a language uses the same expression both as an intensifier and as a reflexive anaphor, this expression is not used as a marker of derived intransitivity (middle marker).» (König and Gast 2006: 236)

[71] It must be noted that Schladt (1999) chose the languages of his sample on the basis of the possibility to identify lexical sources for them, or at least uses characterized by a degree of grammaticalization lower than that of markers of reflexivity. Moreover, the definition of reflexivity on which the author relied to choose the data was rather broad, in order to allow the inclusion of nominal, pronominal and verbal markers in the study (Schladt 1999: 103-104).

that there is no established term in the literature to refer to these expressions, and over the alternative labels already used in the literature ("emphatic reflexives", "emphatics", "reflexives in emphatic use", "emphasizers", "intensive pronouns"; see the relevant references in König and Gast 2006) they prefer the term "intensifier", first used in Moravcsik (1972). This term, despite being sometimes used to indicate grade adverbs such as *very* or *extremely*, is preferred because, according to the authors, it avoids the misleading connotations of the aforementioned terms (König and Gast 2006: 224).[72]

König and Gast describe a number of patterns of variation of intensifiers. Some concern the semantic-pragmatic level, while others concern their morphosyntactic characteristics. Intensifiers can be invariant as Albanian *vetë*, Irish Gaelic *féin* or German *selbst*. They can display gender, number, and case agreement with the focused element – Latin *ips-*, Spanish *mism-*. In some cases, they also have exponents which convey person agreement – English *your-self* '2.POSS-INT', Finnish *itse-nsä* 'INT-3.POSS'. A relationship between intensifiers and reflexives is observed in many languages of the world where the two categories correspond to the same elements: Amharic *ras-*, Arabic *nafs-*, Mandarin Chinese *zìjǐ*, English *him-* / *herself*, Hungarian *mag-*. In other languages, a formal distinction is observed between elements that act as intensifiers and elements that appear in reflexive constructions, see, for example, the following pairs of intensifier and reflexive: Bambara *yɛ̀rɛ̀*/*í*, Bulgarian *sam-*/*se(be)*, Italian *stess-*/*si*, German *selbst*/*sich*. In some cases, however, the reflexive markers have a partial identity with the intensifiers: Georgian *tvit*/*tav-*, Ancient Greek *aut-*/*heaut-*, Hindi *aap*/*apnee aap*, Dutch *zelf*/*zichzelf*.

Puddu (2005) shows that if there is a formal distinction between the two groups, it is actually possible to observe triple series of different strategies, as can be seen in Table 6. In order to complete the typology, to the languages considered by Puddu, we have added English, i.e. a language which does not distinguish between intensifiers and reflexives.

[72] The term intensifier has been proposed by König and his collaborators at least since König (1998) and König and Siemund (1999).

Table 6. Languages that use different forms for reflexives and intensifiers (adapted from Puddu 2005: 49)

Language	Intensifier	Reflexive	"Compounded" reflexive
Italian	*stesso*	*sé*	*sé stesso*
Russian	*sam*	*sebja*	*sam sebja*
German	*selbst*	*sich*	*sich selbst*
English	*himself*		

It can therefore be seen that none of the configurations found in Italian, German and Russian corresponds to a variation in the marker used in English. However, if we proceed beyond reflexive constructions along the grammaticalization cline seen above in (35), we reach an area of functions for which we observe a formal distinction in English as well. Consider, for example, the markers used in verbs of grooming such as 'to wash', or body motion such as 'to sit down' in Table 7.

Table 7. Middle markers

Language	'to wash'	'to sit down'
Italian	*lavar**si***	*seder**si***
Russian	*myt'**sja***	*sadit'**sja***
German	***sich** waschen*	***sich** setzen*
English	*to wash* **Ø**	*to sit* **Ø**

These verbs have a clitic or affixal marker, etymologically linked if not identical to the one used in reflexive constructions, or, as in English, they completely lack a specific mark. The use of these markers – Italian *si*, German *sich*, Russian *-sja* – is not limited to the two groups of verbs exemplified above, but appears, with different language-specific patterns, in other classes of verbs and functions. For an in-depth discussion of this issue, see Haspelmath (2003), which deals with the problems related to the definition of semantic maps, but also with the multifunctionality of middle and reflexive markers. In addressing the latter topic, an indispensable reference can be found in Kemmer (1993), a detailed study of the middle voice from a cross-linguistic point of view. It is nonetheless important to mention here what distinguishes the domain of the middle voice from prototypical reflexivity. To do so, we will now exemplify at least the main functions identified in this area. Figure 12 – issued from Haspelmath (2003) – indicates the main functions of the reflexive and middle domains.

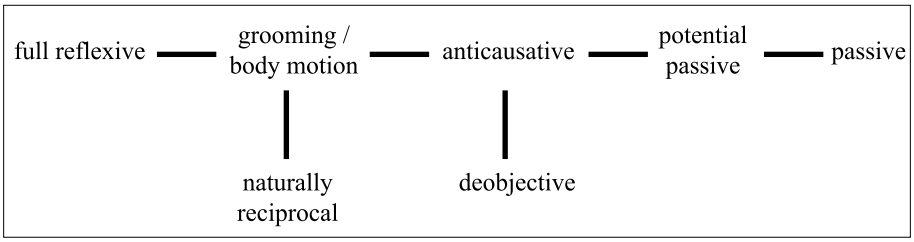


Figure 12. Conceptual space of reflexive and middle functions (Haspelmath 2003: 225)[73]

On one end of the conceptual space, we have prototypical reflexive constructions (*full reflexive*), which are characterized by the coreference of two arguments of a verb that are typically distinct, configuring a generally other-directed action. In this case, the reflexive construction indicates the presence (unexpected, or at least not typical) of a single referent for two arguments of the verb. The action remains transitive, i.e. it clearly distinguishes between two participants in the action, who, however, correspond to a single referent (37).

(37) French: *Judas s'est tué*. (Haspelmath 2003: 223)
'Judas killed himself'

In traditional grammars, verbs of grooming are often regarded as reflexive verbs. In this respect, we can certainly speak of a reflexive situation since those are actions that are actively performed on oneself, but the distinguishability of the participants is lower («low degree of elaboration of events» Kemmer 1993), due to the fact that they are typically directed towards those who perform them – s. (38).a below. A particular case of self-directed action is that of naturally reciprocal actions in which participants in an action act on themselves, and it is not possible to distinguish different transitive actions that are mutually repeated on one and the other participant – s. (38).b.[74] Concerning the verbs of body motion like the one in (38).c, the elab-

[73] Haspelmath (2003) uses the term "semantic map" for both the representation of cross-linguistic generalizations concerning a domain of functional properties, and the language-specific relationship between functions and individual formal strategies. In this book, we will follow Croft (2001b) and we will distinguish between "conceptual space" – defined as «a structured representation of functional structures and their relationships to each other» Croft (2001b: 93) – and "semantic map" – defined as « a map of language-specific categories on the conceptual space» Croft (2001b: 94).

[74] In this regard, compare a sentence like *Gianni e Luca si sono sparati*, which can be interpreted as: a) 'Gianni shot himself and Luca shot himself' (reflexive); b) 'One shot the other', i.e. 'Gianni shot

oration of the event is even lower, and shows how this series of functions is placed on a continuum between actions with two participants (typically transitive actions) and actions with a single participant (typically intransitive actions).

(38) French (Haspelmath 2003: 223)
a. *Bathshéba s'est lavée*
'Batsheba washed'
b. *Elisabeth et Marie se sont rencontrées*
'Elizabeth and Mary met'
c. *Mamoud s'est agenouillé*
'Mamoud kneeled down'

The anticausative function frequently performed by these markers is a method that languages have so as to reduce the number of participants in an action. Typically, this occurs by transforming a transitive action into an intransitive action through the removal of the role of agent, and the maintenance, as an only participant, of the patient of the transitive action – s. (39).a. Alternative to the anticausative, the deobjective function corresponds to the removal of the patient and the maintenance of the agent – s. (39).b.

(39) Russian (Haspelmath 2003: 223)
a. *Dver' otkrylat-s'*
'The door opened'
b. *Sobaka kusaet-sja*
'The dog bites'

Other constructions that imply a reduction in the number of participants have to do with the domain of the passive – potential passive in (40) and passive in (41)[75] – in which, similarly to what happens for the anticausative function, the role

Luca and Luca shot Gianni' (reciprocal). Regardless of the interpretation, in both cases the action of shooting is intended as transitive, and an agent (who shoots) and a patient (who is shot) are clearly identifiable.

[75] The examples provided by Haspelmath for the potential passive are analogous to the argument structures that Kemmer (1993: 147 ff.) calls "facilitative middle". The same construction is also linked by Levin to the "middle alternation" (Levin 1993: 25-26) – *the butcher cuts the meat/the meat cuts easily*: «[The middle construction] is characterized by a lack of specific time reference and by an understood

of agent disappears. However, unlike the anticausative function, which produces intransitive verbs and therefore indicates actions without an agent, the passive has the function of placing the agent in the background, without implying its absence from the action.

(40) German (Haspelmath 2003: 225)
Humboldt liest sich nicht leicht.
'Humboldt doesn't read easily'

(41) Turkish (Haspelmath 2003: 226)

Bu	*iş-ler*	*makine-ler-le*	*yap-ıl-ır*
these	things-PL	machine-PL-INSTR	make-PASS-AOR

'These things are made with machines'

As can be seen from the overview of the functions performed by reflexive and middle markers, this functional domain, in contrast to that of intensifiers, concerns the semantics of the verb phrase, and potentially even the argument structure of verbs as lexemes, regardless of their syntactic use.

For the sake of completeness, it is worth mentioning that a last possible path of grammaticalization that concerns intensifiers leads to the rise of personal pronouns, starting from intensifiers.

4.2 Contrastive coreference markers in contemporary European languages

In this paragraph, we will examine the morphemes that convey the function of contrastive coreference in the European languages considered in the sample described in Chapter 3. We will start from the Indo-European languages (IE) by first examining the groups to which Italian and German belong – Romance and Germanic languages – and we will then move on to consider other groups of

but unexpressed agent. […] These properties distinguish the middle construction from the causative/inchoative alternation. In particular, the intransitive variant of the causative/inchoative alternation, the inchoative [a.k.a. anticausative] construction, need not have an understood agent [and] may have specific time reference […]». (Levin 1993: 26)

the Indo-European family, then the European languages belonging to families different from the IE, and finally, an isolated European language (i.e. Basque).

Although in the previous paragraph we have indicated grammaticalization paths that connect various syntactic functions with the respective lexical sources of origin, detecting the ultimate lexical sources of the morphemes that indicate contrastive coreference in word-formation goes beyond the scope of the discussion of this chapter and this work in general. We will devote a part of Chapter 6 to the discussion of some morphemes that cannot in synchrony be traced back to syntactically free units, and we will consider some possible sources in diachrony. However, our objective is not the exhaustive reconstruction of these grammaticalization paths outside word-formation, but rather the quest for the reasons that have lead a certain element to become a $_{CC}$marker.

4.2.1 IE. Romance languages

As far as the group of Romance languages is concerned, the $_{CC}$marker used in the individual languages always belongs to the group of neoclassical combining forms (see Chapter 2) and corresponds to Italian *auto-* – see (42)-(46).[76]

(42) Italian [GRADIT]
a. ***auto****distruzione*, ***auto****commiserazione*
b. ***auto****distruggersi*, ***auto****commiserarsi*

(43) French [frB1], [frM1]
a. ***auto****destruction*, ***auto****réplication*
b. *s'****auto****détruire*, *s'****auto****répliquer*

(44) Portuguese [ptB1], [ptM1]
a. ***auto****financiamento*, ***auto****-intoxicação*
b. ***auto****financiar-se*, ***auto****intoxicar-se*

[76] Since the spelling of the different markers in the various Romance languages is the same, we will refer to all these Romance morphemes generically with the indication AUTO-. As far as their pronunciation is concerned, the greatest divergence is attested in French, where *auto-* is pronounced [oto-] with the reduction of the diphthong [aw] – as found in Italian, Portuguese and Spanish – to [o]. It should be noted that in Romanian the digraph <au> is generally pronounced as two syllables, and if the diphthong is stressed, then [u] is tonic: e.g. *auto* 'car' [aˈuto] (Beyer *et al.* 1987).

(45) Romanian [roB1], [roM1]
a. ***auto****denunța*, ***auto****dizolvare*
b. *a se* ***auto****denunța*, *a se* ***auto****dizolva*

(46) Spanish [esB1], [esM1]
a. ***auto****castigo*, ***auto****rreplicación*
b. ***auto****castigar se*, ***auto****rreplicar se*

Apart from the common origin in the layer of Greek lexical formants, a syntactically free lexeme corresponding to AUTO- does not exist in any Romance language , if we exclude various clipped forms of Italian *automobile*, French *automobile*, Portuguese *automóvel*, Romanian *automobile*, Spanish *automóvil*. These clippings, however, are second-generation neoclassical compounds (see Chapter 2, section 2.2.5) that also appear to be productive with the meaning 'car': Italian *autostrada* 'motorway', French *auto-école* 'driving school', Portuguese *auto-estrada* 'motorway', Romanian *autobază* 'car depot', Spanish *autobomba* 'car bomb'. In Portuguese and Spanish there is another noun *auto*, derived from Latin *actus* 'act'. It is clear that we can easily exclude any link between *auto-* conveying contrastive coreference, and the nouns *auto* 'car' (a clipping in all Romance languages) or *auto* 'act' (in Portuguese and Spanish).

As in the Italian literature, in French (Mutz 2004) and Spanish (Rainer 1993; Varela and Martin García 1999) AUTO- is treated as a prefix. For Romanian, in Beyrer *et al.* (1987: 59), *auto-* is instead treated as a very productive prefixoid. Note that Beyrer *et al.* (1987) adopt a concept of prefixoid which is very similar to the one explained above in Chapter 2 (section 2.2.6): the behaviour of prefixoids is defined as similar to that of prefixes, which usually do not have a syntactically free correspondent. However, they have a relatively defined and autonomous meaning. Due to their inability to occur freely in syntax, however, they are treated as part of derivation and not of compounding (Beyrer *et al.* 1987: 49).[77]

In Portuguese, compounds with neoclassical combining forms and adjectival compounds with a linking vowel (*lus**o**-brasileira* 'Brasilian Portuguese') are

[77] «Präfixoide oder Pseudopräfixe verhalten sich morphologisch wie Präfixe, d.h., sie kommen in der Regel [...] nicht frei im Satz vor, verfügen aber über eine relativ scharf umrissene, selbständige Bedeutung. Ihrer Syntaktische Unselbständigkeit entsprechend werden sie in unserer Skizze der Wortbildung unter der Ableitung, nicht unter der Komposition, behandelt.» (Beyrer *et al.* 1987: 49)

considered "morphological compounds", as opposed to "morphosyntactic compounds", which do not have linking elements and have properties typical of syntactic and morphological elements (Mira Mateus 2003: 972 ff.). In this classification, the morpheme *auto-* – which is absent from the list of productive prefixes listed in Mira Mateus (2003: 963-64) – is treated as an element of the so-called "morphological compounds", with the final *-o-* intended as a linking element. In this regard, it should be noted that, as a linking element, the *-o-* of *auto-* does not submit to the Portuguese phonological rules, which characterize unstressed vowels within words. It is, for example, expected that a [ɔ] is raised to [u], as in the case of *aut*[u]*rizar* (**aut*[ɔ]*rizar*) 'to authorize', or of lexicalized formations such as *aut*[u]*móvel* (**aut*[ɔ]*móvel*) 'automobile'. Compare the former with *aut*[ɔ]-*sugestão* 'autosuggestion'.

Beyond the different proposals and classifications present in the grammatical traditions of various languages, it can nonetheless be observed that for Romance languages the criteria that have led to considering Italian *auto-* as a prefix – though not a prototypical one – apply. The factors that favor an interpretation of *auto-* as a prefix are above all semantic, with the mismatch between the meaning of the bound element and the possible syntactically free correspondents. On the other hand, the factors that bring AUTO- close to compounding are mainly phonological, and correspond to the fact that AUTO- is non-cohering, bisyllabic and ends in a vowel – characteristics which are not ruled out for prefixes, but which are also typical of many first elements in compounds.

As can be seen from the examples (42)-(46) above, the vocabularies of all these languages record both action nouns and reflexive verbs displaying AUTO-. This shows the compatibility of AUTO- in all Romance languages, as in Italian, with verbs in predicative use. This aspect – which will be taken up later in Chapter 5, section 5.2.1 – leads us to a schematic survey of intensifiers and reflexive markers in Romance languages that is summarized in Table 8 below. [78] The source of the data in the table is the TDIR (with some modifications).[79]

[78] As for the exclusive adverbial intensifier in Italian, the TDIR reports again PRO *stesso*: *la foto l'ho fatta io stesso* 'I took the photo myself'. In Table 3, we exchange the prepositional phrase *da sé* for the exclusive adverbial intensifier: *ho fatto la foto da me* 'I took the photo myself' / *Mario ha fatto la foto da sé* 'Mario made the photo himself'. As a matter of fact, we might consider a sentence like *la foto l'ha fatta Mario/lui stesso* as nothing more than the use of the adnominal intensifier in a sentence with topicalized object.

[79] TDIR is an abbreviation for the *Typological Database of Intensifiers and Reflexives* (Gast *et al.* 2007).

Table 8. Romance languages. Intensifiers and reflexive markers[80]

	Italian	French	Portuguese	Romanian	Spanish
INT adN	NP/PRO *stesso*	(NP) PRO- *même*	NP/PRO *mesmo*	NP/PRO *îns-* PRO.GEN	NP/PRO *mismo*
INT adV	*da sé*	PRO-*même*	PRO *mesmo*	*singur*	PRO *mismo*
INT att	*proprio*	POSS.PRO *propre*	*próprio*	*proprio*	*propio*
REFL H	*sé*	*soi*	*si*	*se*	*sí*
REFL L	*si*	*se*	*se*	*şi*	*se*

The table shows clearly how none of these elements corresponds to the elements involved in $_{CC}$formations, i.e. complex words with the value of contrastive coreference.

The absence of a link with syntactically free elements leads us to exclude an interpretation of AUTO- (as $_{CC}$marker) as an element of compounding.

4.2.2 IE. Germanic languages

Unlike Romance languages, where, as we have seen, $_{CC}$markers appear to be prefixes, in Germanic languages contrastive coreference in word-formation is marked by using markers that can be connected to syntactically free elements that belong to the intensifier-reflexive continuum.[81]

The database is a cross-linguistic collection of examples of intensifiers and reflexives, classified according to the perspective illustrated in König and Gast (2006). When examples from the TDIR are quoted, the original source of the example is also repeated. In most cases, the glosses are provided in the original work or through the TDIR. If a gloss is added or modified in relation to the source, this is noted next to the example.

[80] The abbreviations shown in the first column correspond to the main functions of the markers placed on the intensifiers-reflexive continuum, that is, from top to bottom: INT adN = adnominal intensifier, INT adV = adverbial intensifier (exclusive), INT att = attributive intensifier, REFL H = heavy reflexive, REFL L = light reflexive.

[81] It is worth noting that "syntactically free element" is an accurate term only to a limited extent. On the one hand, different languages may display different constraints on similar word classes, and some languages may have a rather free or, conversely, rigid word order. Yet again, different classes of words in a language may be subject to different types and amounts of constraints as far as their syntactic freedom is concerned. In this sense, more grammaticalized elements (such as intensifiers and especially reflexive markers) should be syntactically more constrained than those that are less grammaticalized (such as nouns).

(47) German: *Selbstzerstörung* 'self-destruction', *Selbstentzündung* 'self-ignition' [DUDEN]

(48) English: *self-destruction*, *self-ignition* [OED]

(49) Dutch: *zelfdoding* 'suicide', *zelfontbranding* 'self-ignition' [nlB1], [nlB2]

(50) Danish: *selvstyre* 'self-government', *selvtænding* 'self-ignition' [dkB1], [dkB2], [dkB3]

(51) Icelandic: *sjálf(s)stjórn* 'self-government', *sjálf(s)kviknun* 'self-ignition' [isB1], [isB2], [isB3], [isM1]

(52) Swedish: *självstympning* 'self-mutilation', *självbevarelse(drift)* '(instinct of) survival, lit. self-preservation' [seB1], [seB2], [seM1]

In (47)-(52), we provide examples of the phenomenon in Germanic languages. The morphemes used in these languages are all cognate with German *selbst*. In spite of the common etymological origin, the syntactically free elements corresponding to the bound elements *self-*, *zelf-*, *selv-*, *sjálf(s)-* and *själv-* belong to different parts of the intensifiers-reflexive continuum and display very different formal features.[82]

We will resume and expand here the discussion of the functions of *selbst* in German, with the objective of using this language as a term of comparison with respect to the other Germanic languages of our sample. In German, intensifiers and reflexives have distinct forms, and *selbst* fulfills both the function of adnominal intensifier (53) and the function of adverbial intensifier (54).[83]

[82] The common ancestor of the morphemes that are used as adnominal intensifiers in Germanic languages can be traced back to the Proto-Germanic **selba* (POKORNY: 884). The German form *selbst* evolves from *sëlpes*, the genitive form of the Middle High German identity pronoun *sëlp*, to which an epenthetic *t* is added in Late Middle High German (Paul 1998 [1881]: §222).

[83] «[A] feature that is typical of SAE languages is the differentiation of reflexive pronouns and intensifiers» (Haspelmath 2001: 1501). See also Chapter 3, section 3.2.1.

(53) *der Direktor selbst wird mit uns sprechen.* (TDIR)
ART director INT will with us speak
'The director himself will talk to us.'

(54) *Olga unterrichtet ihre Kinder selbst.* (TDIR)
Olga teaches her children INT
'Olga teaches her children herself.'

Similarly to German, Danish and Swedish use the markers *selv* and *själv,* respectively. In the following examples, we see the use of the adnominal – in (55) – and the adverbial intensifier – in (56) – and in (57), the reflexive used with a typically other-directed verb in Danish.

(55) *direktør-en ham selv vil tale med os.* (TDIR)
director-ART INT INT will talk with us
'The director himself will talk to us.'

(56) *Santiago dræbte selv fisk-e.* (TDIR)
Santiago killed INT fish-ART
'Santiago killed the fish himself.'

(57) *Jan taler med sig selv.* (TDIR)
Jan talks with REFL INT
'Jan is talking to himself.'

It is worth noting that *selv*, like German *selbst*, is invariable. Swedish *själv*, on the contrary, agrees with the pronoun or the noun phrase it refers to in gender and number: while in (58) we have the definite (weak) form *själva*, in (59) and in (60) we find *själv* in the singular indefinite (strong) form of the common gender, which has no agreement exponents.

Here below in (58), we have an example of the adnominal intensifer, in (59) of the adverbial exclusive, and in (60) of the heavy reflexive in Swedish.

(58) *själv-a* *kung-en* *har sovit här*. (Holmes and Hinchliffe 1994: 147; gloss: TDIR)
INT-DEF king-ART has slept here
'The king himself has slept here.'

(59) *jag* *kan göra* *det* *själv*. (Holmes and Hinchliffe 1994: 146; gloss: TDIR)
I can do that INT
'I can do that myself.'

(60) *han* *hjälper* *bara sig* *själv*. (Holmes and Hinchliffe 1994: 147; gloss: TDIR)
he helps only REFL INT
'He helps only himself.'

Icelandic displays the richest inflectional inventory. In this language, the adnominal and the adverbial exclusive intensifier agrees with the controller of agreement in gender, number and also in case – examples (61)-(62), respectively.

(61) *verkalýðshreyfingin* *sjálf* (Kress 1982: 173; gloss: TDIR with modifications)
the.labour.movement INT.F.NOM
'The labour movement itself'

(62) *ég* *gerði* *thað* *sjálfur* ([isB3])
I did that INT.M.SG.NOM
'I did it myself/alone'[84]

In Icelandic as well, the reflexive pronoun has case marking (63), and, to the best of our knowledge, it does not need to be combined with the intensifier like the other Germanic languages seen so far.

(63) *hún* *skammast* *sín* (Kress 1982: 173; gloss: TDIR with modifications)
he is.ashamed.of REFL.GEN
'He is ashamed of himself.'

[84] The translation of the example in the Icelandic-Italian dictionary ([isB3]) is 'L'ho fatto io stesso/da solo'.

By clearly differentiating between reflexives and intensifiers, Icelandic constitutes an extreme case along the range of variation of Germanic languages. Closer to the opposite extreme of languages in which intensifiers and reflexives are not formally distinguished there are two West Germanic languages . Dutch, on the one hand, shows a partial identity between the two series of markers, since *zelf* always accompanies the pronoun (*sig* in the third person) in heavy reflexives – in contrast with German in which *selbst* may be omitted.

(64) *de universiteit zelf bereik je met de bus.* (TDIR)
ART university INT reach you with ART bus
'The university itself is reached by bus.'

(65) *mijn post maak ik altijd zelf open.* (TDIR)
my letters make I always INT open
'I always open my letters myself.'

(66) *Jan haatte zichzelf.* (TDIR)
Jan hated REFL
Jan hated himself.

English constitutes the extreme case of a language in which intensifiers and reflexives are not formally distinguished.

(67) *the director himself will talk to us.* (TDIR)
ART director INT FUT talk to us
'The director himself will talk to us.'

(68) *I have swept the court myself.* (TDIR)
I have swept ART court INT
'I have swept the court myself.'

(69) *Paul saw himself in the mirror.* (TDIR)
Paul saw REFL in ART mirror
'Paul saw himself in the mirror.'

It is beyond the scope of this overview to cover the uses of compounded reflexive markers – i.e. a reflexive pronoun combined with an intensifier – and those of simple reflexive pronouns. Here, with a rough distinction, we will consider compound markers as heavy reflexives – i.e. coreference markers used with verbs indicating typically other-directed actions – and simple markers as light reflexives – i.e. markers indicating coreference in the case of typically self-directed actions. Also in this case, Icelandic and English constitute two extremes: the first because it uses the reflexive pronoun both as a heavy and as a light reflexive marker; the second because it completely lacks a light reflective marker – compare the English translation of the Icelandic example (70).

(70) *Jón rakar sig* (Sigurjónsdóttir and Hyams 1992: 369; our gloss)
John shave.IND.PRES.3SG REFL.NOM.SG
'John shaves'

A schematic survey of intensifiers and reflexive markers in Germanic languages is summarized in Table 9 below.

Table 9. Germanic languages. Intensifiers and reflexive markers

	German	English	Dutch	Danish	Icelandic	Swedish
INT adN	NP/PRO *selbst*	NP PRO-*self*	NP/PRO *zelf*	NP/PRO *selv*	NP/PRO *sjálf-*	NP/PRO *själv-*
INT adV	*selbst*	PRO-*self*	*zelf*	*selv*	*sjálfur*	*själv*
INT att	*eigen*	own	*eigen*	*egen*	*eigin*	*egen*
REFL H	*sich (selbst)*	PRO-*self*	*zichzelf*	*sig (selv)*	*sig*	*sig (själv)*
REFL L	*sich*	–	*zich*	*sig*	*sig*	*sig*

The overview of the uses of intensifiers and reflexives in Germanic languages allows us to observe some tendencies as far as the differences between word-formation and syntactic markers are concerned. First, as already mentioned above, all Germanic languages use the outcomes of Proto-Germanic **selba* as $_{CC}$markers. The syntactically free units with which they can be linked belong to the intensifiers-reflexive continuum, and occur in formally different syntactic constructions, which nonetheless share at least the function of adnominal and adverbial exclusive

intensifiers. Excluding Icelandic, in the other Germanic languages these elements are also involved in the expression of reflexivity when verbs indicating typically other-directed actions are used.

We have mentioned above that in different Germanic languages intensifiers display different inflectional features: some are invariable (German *selbst*, Dutch *zelf*, Danish *selv*), some inflect and agree with the focused element in gender, number (Swedish), and case (Icelandic). English configures a special case in which the compound intensifier/reflexive displays a paradigm of forms inflected for gender number and person. In cases of invariable elements, the bound form corresponds to a syntactically free unit. In the cases of inflectionally variable markers, the bound forms correspond to one of the forms of the paradigm. In Swedish, the bound form corresponds to the singular indefinite form of the common gender, i.e. *själv*, that is, the form of the paradigm without inflectional exponents. In Icelandic, two forms can occur, *sjálf-* or *sjálfs-*, corresponding respectively to the base to which the inflectional exponents are concatenated, and to the form of the masculine and neuter genitive. English is again a special case, because the bound element *self-* used in word-formation is a bound morpheme in syntax as well, as it is compounded with a pronominal form which precedes it.[85]

The elements exploited in the formations, as seen above in (47)-(52), can therefore be considered units in a compounding process, but as their corresponding (syntactically) free units, they have highly grammaticalized forms that occur in specific constructions in which they occupy syntactic positions that are sometimes determined in a rather rigid way: see, for example, the relative position of the intensifier with respect to the element it focuses. These features of the sources of $_{CC}$markers show that they do not belong to one of the major word classes.

As in the case of *selbst*, it should be borne in mind that in these languages there are nouns indicating 'the ego, the self' that could provide support, at least formally, for the use of these elements in compounding. A particularly significant case is that of the English intensifier/reflexive X-*self*: the element *self* occurs as a syntactically free unit only in its nominal use (*the self*). Nouns corresponding to the intensifiers that have this same lexical meaning are also present in the other Germanic languages: compare Dutch *zelf*, Danish *selv*, Icelandic *sjálf*, Swedish *själf*. None of these nouns, however, is connected to the values of coreference or

[85] The paradigm of X-*self* consists of the univerbation of a form attributable to the possessive adjective (*my-*, *your-*, *herself*, *your-*, *ourselves*) and the bound element *-self*/*-selves*. In some forms, however, the first element corresponds to the oblique forms of the personal pronoun (*him-*, *itself*, *themselves*).

contrastivity that the bound word-formation markers convey in compounds.

Germanic languages also behave homogeneously with respect to the possibility of having elements that serve as paradigmatic alternatives to those examined so far. If in these languages an element different from an outcome of Proto-Germanic **selba* is present in $_{CC}$formations, then the marker used always corresponds to the marker that fulfills the function of attributive intensifier in syntax. For the attributive intensifier, Germanic languages have cognate terms, as can be seen in Table 9 above. The attributive intensifier therefore acts as a secondary $_{CC}$marker, and can occur in various formations: German *Eigen·transplantation* 'autotransplantation', English *own goal* 'autogol', Swedish *egen·företagare* 'self-employed', Dutch *eigen·liefde* 'self-love', Danish *egen·navn* 'proper name'. In addition to the attributive intensifier *eigin*, Icelandic also uses another element in word-formation, *sér-,* a homograph of the dative of the reflexive pronoun *sig*: *sér·navn* [isB2] 'proper name [Eigenname]', *sér·vitringur* [isB2] 'asocial [Eigenbrötlerisch]'.

4.2.3 IE. Slavic languages

Slavic languages are rather homogeneous as far as intensifiers and reflexives are concerned. In all Slavic languages here considered, the adnominal (71) and the adverbial exclusive intensifier (72) are expressed using *sam*, an element which, like adjectives, displays gender, number and case agreement with its controller. In this last function, *sam* can mean 'alone', and in some cases it is a synonym for 'solitary (German *einsam*)', for example, in Polish, where it is indicated as a synonym for *samotny* 'solitary'.[86]

(71)

a. Croatian (Batinić Angster 2020: 114; our gloss)

On	*je*	*izabrao*	*da*	*siđe*	*u*
3SG.M.NOM	be.PRS.3SG	choose.PART.M.SG	that	go.down.PRS.3SG	in
svoj	*svijet,*	*da*	*sam*	*sebe*	*ograniči*
POSS.REFL.M.SG.ACC	world.SG.ACC	that	INT.M.SG.NOM	REFL.ACC	limit.PRS.3SG

'He decided to sink into his own world, which he himself delimitates'

[86] On the relation of lexemes meaning 'alone' – such as *sam* in Slavic languages – with intensifiers, also compare Italian *da solo* or German *allein* as exclusive adverbial intensifiers.

b. Czech (Naughton and von Kunes 2021: 127; our gloss)
Napsal *to* *sám* *prezident*
wrote.M.SG that INT.M.SG.NOM president(M).SG.NOM
'The president himself wrote it'

c. Polish (TDIR; with modifications)
dyrektor *sam* *będzie* *z* *nami* *rozmawiał*
director(M).SG.NOM INT.M.SG.NOM will with ust alk.PART.M.SG
'The director himself will talk to us'

d. Russian (TDIR; with modifications)
direktor *sam* *budet* *govorit'* *s* *nami*
director(M).SG.NOM INT.M.SG.NOM will speak with us
'The director himself will talk to us'

(72)

a. Croatian (Batinić Angster 2020: 114; our gloss)
Iznenadila *sam* *i* *sam-u* *sebe*
surprise.PART.F.SG be.PRS.1SG also INT-F.SG.ACC REFL.ACC
'I surprised even myself'

b. Czech (Naughton and von Kunes 2021: 127; our gloss)
Napsala *to* *celé* *sama*
wrote.F.SG that all.N.SG.ACC INT.F.SG.NOM
'She wrote it all herself'

c. Polish (TDIR; with modifications)
Olga *uczy* *swoje* *dzieci* *sama*
Olga teaches POSS.REFL child.ACC.PL INT.F.SG.NOM
'Olga teaches her children herself'

d. Russian (TDIR; with modifications)
moi *pisma* *ja* *vsegda* *raspečatyvaju* *sam/a*
my letters I always I.open INT.SG.M/F
'I always open my letters myself'

As for reflexive markers, beyond the obvious possible phonological and inflectional peculiarities, Slavic languages display cognate reflexive pronouns, neutralized for person and number agreement, which inflect for case. As for all other pronouns, two series are available, one stressed (heavy reflexive (73)) and another unstressed (light reflexive (74)).

(73)

a. Croatian (Batinić Angster 2020: 67; our gloss)
Studenti *govore* *o* *sebi*
student.PL.NOM speak.PRS.3PL about REFL.LOC
'The students talk about themselves'

b. Czech (Naughton and von Kunes 2021: 99; our gloss)
Často *se* *na* *sebe* *dívá* *do* *zrcadla*
often REFL at REFL.ACC looks in mirror.SG.GEN
' S/he often looks at her/himself in the mirror'

c. Polish (Skibicki 2007: 175; our gloss)
Widział *siebie* *w* *lustrze*
saw.M.SG REFL.ACC in mirror
'He saw himself in the mirror'

d. Russian (Mulisch 1975: 328; gloss TDIR)
Aleksej *mnogo* *rasskazal* *o* *sebe*
Aleksei much told about REFL.DAT
'Aleksei has talked a lot about himself'

(74)

a. Croatian (Batinić Angster 2020: 64; our gloss)
Ivan *se* *ubio*
Ivan.SG.NOM REFL.ACC kill.PTCP.M.SG
'Ivan killed himself'

b. Czech (czB2; our gloss)
dobře *se* *obléká*
good REFL.ACC dress.PRS.3SG
'S/he dresses good'

c. Polish (TDIR; with modifications)
on *kocha* *sie* *sam* *ponad wszystko*
He loves REFL.ACC INT.M.SG.NOM above all
'Most of all he loves himself.'

It must be noted that the accusative form of the stressed reflexive in Croatian is *sebe* (see also example (73).a above), and in Russian is *sebjà*. The Russian light reflexive developed as an affix, *-sja* (see Table 7 above and the examples of intransitive verbs under (39)).

Beside the reflexive pronoun, Slavic languages also have the reflexive possessive (75). They share cognate markers in this respect as well.

(75)

a. Croatian (Batinić Angster 2020: 64; our gloss)
Marin *je* *ugledao* *svoj-u* *mačk-u*
Marin.SG.NOM be.PRS.3SG sight.PTCP.M.SG POSS.REFL-F.SG.ACC cat(F)-SG.ACC
'Marin saw his cat'

b. Czech (Naughton and von Kunes 2021: 110; our gloss)
Karel *ztratil* *svoje* *pero*
Karel has.lost.M.SG REFL.POSS.N.SG.ACC pen
'Karel has lost his [own] pen.'

c. Polish (Skibicki 2007: 179; our gloss)
Oni *czytają* *swoją* *książkę*
they read.PRS.3PL POSS.REFL.F.ACC book.ACC
'They read their (own) book'

d. Russian (Stanković 1998; our gloss)
ljubiti *svoju* *rodin-u*
love.INF REFL.POSS homeland-SG.ACC
'to love (one's) own homeland'

The Czech possessive reflexive, as all other possessives, has both a long (*svoje* in (75).b) and a short form (*své*). We will see below that the latter is used in word-formation.

It is in the domain of attributive intensifiers (76) that Slavic languages show some divergence. In (76).b and (76).e, we can observe the existence of a second marker, which in Serbian competes with *vlastit* (compare Croatian in (76).a).[87]

(76)

a. Croatian (Batinić Angster 2020: 64; our gloss)
Marin *je* *ugledao* *vlastit-u* *mačk-u*
Marin.SG.NOM be.PRS.3SG sight.PTCP.M.SG own-F.SG.ACC cat(F)-SG.ACC
'Marin saw his own cat'

b. Serbian (Stanković 1998; our gloss)
po *sopstvenoj* *želji*
according.to own.F.SG.DAT desire(F).SG.DAT
'according to (one's) own desire'

c. Czech (Naughton and von Kunes 2021: 111; our gloss)
Mluvím *o* *svém* *vlastním* *bratrovi*
I.talk about POSS.REFL.M.SG.DAT own.M.SG.DAT brother.SG.DAT
'I'm talking about my own brother.'

[87] The Russian attributive intensifier *sobstvennyj* and the Serbian *sopstven* are not etymologically related to those in other West and South Slavic languages. In Russian, *sobstvennij* is linked to the noun *sobstvo* 'property, quality', but it has actually developed from the Old East Slavic noun of the same meaning *sobьstvo* (Vasmer 1964-1973). In Serbian, the adjective, despite appearing as derived (compare *zdravstvo* 'health' ~ *zdravstveni* 'related to health'), does not appear to be linked to any base in synchrony.

d. Polish (plB4; our gloss)

znać	*coś*	*jak*	*własn-ą*	*kieszeń*
know.INF	something	as	own-F.SG.ACC	pocket.ACC

'to know something like (one's) own pocket'

e. Russian (Stanković 1998; our gloss)

po	*sobstvennomu*	*želannju*
according.to	own.M.SG.DAT	desire.SG.DAT

'according to (one's) own desire'

Turning to word-formation,[88] in all Slavic languages compounds with a $_{CC}$marker have the structure *sam-* followed by the linking vowel *-o-* and by the second element of compounding.[89] The morpheme *sam-* can be traced back to the adnominal and the exclusive adverbial intensifier – see Table 10.

Table 10. Comparison of $_{CC}$markers in Slavic languages

Croatian	Serbian	Czech	Polish	Russian	
***sam**·o·obrana*	***sam**·o·odbrana*	***sebe**·obrana*	***sam**·o·obrona*	***sam**·o·obrona*	'self-defence'
***sam**·o·kontrola*		***sebe**·kontrola*	***sam**·o·kontrola*	***sam**·o·kontrol'*	'self-control'
***sam**·o·ljublje/**sebe**·ljublje*		***sebe**·láska*	***sam**·o·lubstwo*	***sebja**·ljubje*	'self-love/ selfishness'
***sam**·o·ubojstvo*	***sam**·o·ubistvo*	***sam**·o·vražda*	***sam**·o·bójstwo*	***sam**·o·ubijstvo*	'suicide'
***vlast**·o·ručan*	***svoje**·ručan*	***vlastn**·o·ruční*	***własn**·o·ręczny*	***sobstvenn**·o·ručnyj*	'by one's own hand'
***svoje**·glav*		***své**·hlavý*	–	***svoe**·nravnyj*	'stubborn'

In addition to *sam-*, other $_{CC}$markers can be observed in Slavic languages. In some formations, it is possible to find a morpheme comparable to the stressed

[88] A series of lexicographic sources has been consulted to exemplify $_{CC}$formations in Slavic languages: Croatian (hrB1, hrB2, hrB3), Serbian (srB1, srB2), Czech (czB1, czB2), Polish (plB1, plB2, plB3, plB4), Russian (ruB1, ruB2, ruB3, ruB4).

[89] In Slavic languages, the most widely used linking element in compounds is *-o-*, but it is possible to have the presence of other linking vowels – see Russian *-e-* in *ogn-e-tušitel* 'fire extinguisher' (Benigni and Masini 2009: 174) – or of no linking element – such as compounds containing *mistrz* in Polish: *baler-mistrz* 'dance master' (Szymanek 2009: 467).

form of the reflexive pronoun. Czech in particular uses the marker *sebe-* in many compounds where other Slavic languages present *sam-* instead.[90]

As in German and most Germanic languages, secondary markers are available in Slavic languages as well. These secondary markers correspond either to the attributive intensifier (Croatian and Serbian *vlast-*, Czech *vlastn-*, Polish *własn-*, Russian *sobstvenn-*) or to the reflexive possessive (Croatian and Serbian *svoje-*, Czech *své-*, Russian *svoe-*). As we will see later in paragraph 5.2.3, these markers have functions similar to the ones fulfilled by the German marker *eigen-*.

To summarize the case of the Slavic languages considered here, we can observe a substantial homogeneity concerning the main marker SAM-. As for the markers alternative to SAM-, the situation becomes more complex, given the presence of at least three other types of lexical sources – the reflexive pronoun, the reflexive possessive, the attributive intensifier. Moreover, four different roots are involved: beyond the cognates of Croatian and Serbian *sam*, *svoj* and Croatian *vlastiti*, the Serbian and Russian attributive intensifiers *sopstveni* and *sobstvennyj*, respectively.

4.2.4 IE. Celtic languages

We will now consider the last two groups of the Indo-European languages of Europe: Celtic and Baltic languages. With respect to the lexical sources used in complex formations to convey contrastive coreference, the languages of these two groups display a less homogeneous picture compared to that seen so far for Romance, Germanic and Slavic languages.

In Celtic languages, we can observe two types of strategies. The first type is compositional, and involves three different morphemes in three Celtic languages. The first morpheme we consider is *féin-*, present in Irish Gaelic. A corresponding non-bound morpheme, *féin*, acts in Irish Gaelic as the adnominal intensifier (77), the adverbial (78), as reflexive (79), and as attributive intensifier (80).

[90] Also in Serbian and Croatian, it is possible to have *sebe-* in compounds. For example, the pair *samoljublje*/*sebeljublje* (Herman 1975: 255) is attested along with Czech *samolibost*/*sebeláska* [czB2] or Russian *samoljubje*/*sebjaljubje* [ruB2].

(77) (TDIR)
fanóidh mise mé féin leis na páistí.
will.stay I.EMPH I INT with.them ART children
'I myself will stay with the kids.'

(78) (TDIR)
osclaím féin mo chuid litreacha i gcónaí.
open.1SG INT my portion letter.GEN.PL always
'I always open my letters myself.'

(79) (TDIR)
is minic Max ag caint leis féin.
is often Max at talking with.[him INT]=REFL
'Max often talks to himself.'

(80) (TDIR; slightly modified gloss)
tá m'eochair féin agam.
is my.key INT at.me
'I have my own key.'

In Scottish Gaelic, the functions of intensifier (81) and reflexive (82) are performed by *fhèin*, cognate with the Irish Gaelic marker.

(81) ([scB1]; our gloss)
Chan eil clan agam fhèin.
not is issue to.me INT
'I myself have no children' ['Ich selbst habe keine Kinder.']

(82) (MacAulay 1992b: 178)
Tha Iain ga mharbhadh fhèin.
is Iain at.his killing.VN self
'Iain is killing himself.'

Finally, in Welsh there is a different element, *hun*, *hunan* (plural *hunain*), which just like *féin*/*fhèin* acts as reflexive (83), intensifier (84), and can also occur as

attributive intensifier in the possessive construction (85).

(83) (Thomas 1992: 293)
Mae Mair wedi lladd ei hun(an).
is Mair in kill her self
'Mair has killed herself'

(84) (Thomas 1992: *ibidem*)
Mae'r gweinidog ei hun wedi galw.
is-the minister his self after call
'The minister himself called'

(85) (Thomas 1992: *ibidem*)
Mae Mair wedi malu ei char ei hun.
is Mair after smash her car her self (i.e. 'of her self')
'Mair smashed her own car'

A situation similar to that of Welsh is attested in Breton, where a cognate of Welsh *hun*, *hunan*, is also linked to the numeral *unan* 'one' in Breton. Breton *–(h) unan* is always postponed to a possessive affix, and can act as the adnominal intensifier in a reflexive function (86), as adverbial intensifier (87) and as attributive intensifier (88).

(86) ([brB1]; our gloss)
komz outañ e-unan
speak to.3SG.M 3.SG.M-INT
'to speak to oneself' [zu sich selbst sprechen]

(87) ([brB1]; our gloss)
klevet em eus an dra-se me va-unan
heard me is (=I have) DEF thing-that 1.SG 1.SG-INT
'I have heard it myself' [ich habe es selbst gehört]

(88) (Press 1986: 100; our gloss)
em *falez* *va-unan*
in.1SG palace 1.SG-INT
'in my own palace'

Interestingly, even though the case of Breton parallels that of the other Celtic languages, especially that of Welsh, the intensifier is not involved in word-formation (see further below).

Irish Gaelic also displays another $_{CC}$marker that is apparently only bound, *ùath-*, which means 'only, mono-, self-' (see [irB1] and [irB2]). Some examples are given below in (89) to illustrate these strategies.

(89) Celtic languages: compouding
Irish Gaelic: a. *féin·diúltú* 'self-denial', *féin·adhaint* 'self-ignition' [irB1], [irB2]
Irish Gaelic: b. *uath·ghiniúint* 'self-generation', *uath·adhaint* 'self-ignition'
Scottish Gaelic: *féin-riaghladh* 'self-government' [scB1]
Welsh: *hunan·leiddiad* 'suicide', *hunan·ymholiad* 'self-evaluation' [cyM1]

A second type of strategy observed in Welsh and Breton does not belong to compounding but to derivation. It involves a preverb having locative value ('around'), which has also developed reciprocal or reflexive value. It is worth noting that this strategy is primarily a word-formation process involving the production of derived verbs, which in turn give rise to deverbal nouns.

The formations with this type of morpheme are few, and among them we have both formations in which the contribution of the marker is clear and akin to that of other $_{CC}$markers, and idiomatized formations where the contribution of the different elements is less clearly identifiable: see, for example, Welsh *ym·ladd* 'battle, ≈ killing each other', Breton *em·gann* 'battle, ≈ fighting each other'). In Welsh, the formations with *ym-* appear particularly idiomatic, and closer to that of affixal markers of anticausativity or of light reflexive markers. In Breton, the productivity of *em-* is also low, but transparent formations characterized by contrastive coreference can be recognized: *em·gar* 'selfishness, ≈ self-love', *em·skiantek* 'self-awareness', *em·zifenn* 'self-defense' ([brB1]).[91]

[91] On Breton *em-*, its cognates in Celtic and its etymology see also Chapter 6, section 6.1.2.

The case of Breton is rather problematic since the reflexive can also be expressed by using the marker *en em*, defined in the literature as a leniting particle (Press 1998: 162; Desbordes 1990: 68), which has both reflexive and reciprocal meaning.

(90) (Press 1986: 162; our gloss)
ar vugale en em vag
DEF children REFL feed
'The children feed themselves'

(91) (Press 1986: 163; our gloss)
an armeoù en em gann
DEF army.PL REC fight
'The armies fight'

The range of functions of this marker include, beside reflexive and reciprocal, middle marking, and more idiomatized cases in which the marker is actually just part of the phonological make-up of an otherwise unitary meaning (see Favereau 1997: 264; Press 1986: 163; Desbordes 1990: 68).

One might consider Breton *en em* as a light reflexive marker, similar to Italian *si* or French *se*. It should not be ignored that *en* is actually a 3rd person pronoun, which, in the light reflexive construction, is generalized to all persons in contemporary Breton (see Favereau 1997: 263), while *em* is originally a verbal prefix. Despite being considered together with *en* as a particle and no longer as a prefix, *em* precedes the lexical verb, and is not preposed to the auxiliary (Press 1998: 162). We thus consider it a prefix, and as such, it also occurs in the light reflexive construction. The perception of *en em* as a unit and as a somewhat independent particle is justified by the fact that it is often attested in reduced forms like *'n em*, *'nom*, *'non*, *'nim* (Favereau 1997: 263), and it might constitute a case of degrammaticalization of a reflexive/reciprocal prefix into a (complex) reflexive/reciprocal clitic.

4.2.5 IE. Baltic languages

As for Celtic languages, Latvian and Lithuanian, despite being genetically very closely related languages, they differ in terms of the morpheme used to convey contrastive coreference in word-formation.

Latvian aligns itself with Germanic, Slavic and (partly) Celtic languages by using the adnominal intensifier (92). The same marker *pats* is also involved in the exclusive adverbial intensifier (93).

(92) ([lvB1]; our gloss)

es	*pats*	*to*	*padarīšu.*
I	INT.M.SG	DEM.SG.ACC	will.do

'I myself will do it [ich werde es selbst (selber) tun]'

(93) ([lvB1]; our gloss)

to	*viņš*	*darīja*	*pats*	*no*	*sevis*.
DEM.SG.ACC	3SG.NOM	did	INT.M.SG	from	REFL.SG.GEN

'He did it by himself [das tat er von selbst]'

Note that the pronoun *pats* – also called "identity pronoun" – has a rich paradigm of forms that inflect for case, gender and number. The citation form *pat-s* corresponds to the masculine singular nominative, and can be analyzed as the union of the base *pat-* with the suffix *-s*. If we exclude the feminine singular nominative *pat-i*, the other forms of the paradigm are obtained by juxtaposing the alternative base *paš-* with the relevant suffixes.

As can be seen from example (94), this second base of the intensifier *pats* appears as a modifier in compounds.[92]

(94) *paš kontrole* 'self-control; *paš·darināts (alus)* 'home-brewed beer' [lvB1]

Although it has a similar mark to Latvian in the function of adnominal (95) and adverbial intensifiers (96), Lithuanian draws upon the paradigm of the re-

[92] In a few formations, the base used in composition is instead *pat-*: e.g. *pat·mīlība* 'self-love (German *Selbstliebe*)', *pat·valdība* 'despotism (German *Selbstherrschaft*)'.

flexive pronoun *savęs* (97) to convey contrastive coreference in word-formation.[93] In Table 11, we reproduce the complete paradigm of the Lithuanian reflexive pronoun.

Table 11. Paradigm of the Lithuanian reflexive pronoun

nominative	–
genitive	*savęs*
dative	*sáu*
accusative	*savè*
instrumental	*savimì*
locative	*savyjè*

The reflexive pronoun lacks the nominative and, due to the fact that its antecedents can be participants of any person and number, it also lacks the plural (Ambrazas 1997: 192).

(95) (TDIR)
mùs priims pàts direktorius.
us will.welcome INT.M.SG.NOM director
'The director himself will welcome us.'

(96) (TDIR, with modifications)
man skirtus laiškus atplėšiu visuomet aš pàts.
me allot.PTCP letters tear.open always I INT.M.SG.NOM
'I always open myself the letters addressed to me.'

(97) (Senn 1966: 189; gloss: TDIR)
jìs perka sáu knỹgą.
he buy REFL.DAT book
'He is buying himself a book'

From the examples proposed in (98), it can be seen that the bound marker *savi-* alternates with another marker: *sava-*. It is probable that the two morphemes must

[93] An example of the use of the cognate Latvian reflexive pronoun can be seen in (53) above.

be traced back to the root *sav-*. This root also occurs in various forms of possessive pronoun: the genitive *sàvo*, the possessive adjectives *sãvas*, *-à* and the more common *savàsis*, *-óji* (Ambrazas 1997: 193). Two different linking vowels complete the form of the two markers: *-i-* (typical of nouns with stem in *-i-* Mathiassen 1996: 57) and *-a-* (typical of nouns with stem in *-o-* and of a part of the adjectives Mathiassen 1996: 57-71). In this regard, it should be noted that reflexive pronoun (like other personal pronouns not inflected by gender) has instrumental and locative endings similar to those of nouns with stem in *-i-* (*-imi*, *-yje*; Ambrazas 1997: 123, 208), while for the remaining cases, it has endings specific of personal pronouns (Ambrazas 1997: 207-208).

(98) a. *savi·tvarda* 'self-control [Selbstbeherrschung]' [ltB2], *savi·indukcija* 'self-induction' [ltB1]
b. *sava·rankiškas* 'autonomous [eigenhändig]', *sava·naudiškas* 'self-interested [eigennützig]' [ltB2]

Stundžia (2016: 3092-3093) explicitly connects *sav-* with the interfix *-i-,* as in *sav-ì-gyn-a* 'self-defence' to *sav-ė* 'himself, herself', i.e. the reflexive pronoun, and links *sav-* with the interfix *-a-,* as in *sav-a-nõr-is*, *-ė* 'volunteer' to *sãv-as*, *-à* 'one's own', i.e. the reflexive possessive. Consequently, the change of a linking element in Lithuanian is the only formal device distinguishing formations that in German – as can be seen by the German translations of the Lithuanian formation in (98) – would be expressed by *selbst-* and *eigen-,* respectively.

4.2.6 Other IE. languages

In this section, we will consider two Indo-European languages which do not belong to any wider group of the Indo-European languages: Modern Greek and Albanian.

4.2.6.1 Modern Greek

The situation of Modern Greek is noteworthy. In compounds, this language has a marker that is similar to that found in Romance languages, but is not involved in the reflexive intensifier continuum (99).

(99) *afto·sevasmós* 'self-respect [rispetto di sé]' [elB1][94] *afto·apaskhólisi* 'self-employment [lavoro autonomo]' [elM1]

In fact, the adnominal intensifier in Greek is today expressed with the construction NP / PRO ART *ídhi-*, as can be seen in the example (100).

(100) (Joseph and Philippaki-Warburton 1987: 161; gloss: TDIR with modifications)

eghó	*o*	*ídhios*	*to*	*ékana*
I.NOM	the.NOM.SG.M	INT.M.SG.NOM	it	did

'I myself did it.'

The Modern Greek emphatic pronoun *aftó* (*αυτός*) evolved from the Classical Greek pronoun *autós* 'same', once the adnominal intensifier, from which the $_{CC}$marker *afto-* also developed. The Modern Greek *aftó* is part of the series of strong forms of personal pronouns which are used «when emphasis or distinction is required» (Holton *et al.* 2012: 113). An example of this emphatic use, which can be compared to that of the adnominal intensifier, is presented in (101).

(101) (Holton *et al.* 2012: 390; our transliteration and gloss)

aftós/afté	*to*	*ékane*
3SG.M/3SG.F	it	did

'He/she [and no one else] did it.'

Today, the marker of reflexivity in Modern Greek consists of *eaftós* (*εαυτός*) – via its interpretation as a singular noun: 'the self' – preceded by the definite article and followed by a possessive enclitic pronoun inflected for person and number (102).[95]

[94] The right constituent of the formation derives from the verb *sévome* / *σέβομαι* 'respect [rispettare]'.

[95] The history of reflexive pronouns from Homeric Greek to Modern Greek is very complex. Puddu (2010: 392-397) chronicles it, showing, first of all, how the reflexive pronoun has displayed, at least in the third person, a compound form (*hé autón*) as early as Homeric Greek. In this period, reflexivization strategies differ between third person and first and second person, where an anaphoric pronoun is used, possibly reinforced by the intensifier. In the outline of subsequent developments, first we can see a trend towards the adoption of a regular reflexivization strategy, organized according to a paradigm similar to that of anaphoric personal pronouns; subsequently, a second trend leads to the progressive flattening of all persons (first those of the plural in the post-classical era, then also those of the singular in the Byzantine era) on a single form (*heaut-*). In addition to the dynamics of paradigm restructuring,

(102) (Holton *et al.* 2012: 394; our transliteration and gloss)

prépi	*na*	*frontísume*	[*ton eaftó mas*]
it.is.necessary	to	look.after.DEP.1PL	REFL.1PL.ACC

'We must look after ourselves'

In Modern Greek, therefore, an element that in its syntactically free use does not belong – at least synchronically – to the intensifiers-reflective continuum is used as $_{CC}$marker. The case of (Modern) Greek exemplifies the development of personal pronouns from intensifiers, a phenomenon of grammaticalization which constitutes a path diverging from the intensifiers-reflexive continuum.

4.2.6.2 Albanian

The last language of the Indo-European family considered in this study is Albanian.[96] In this language, the element that conveys contrastive coreference is *vet(ë)-* (103).

(103) *vetë·rrjedhje* '(liquid) discharge, secretion';[97] *vet·asgjësim* 'self-destruction', litt. self-annihilation',[98] [sqB2]

The syntactically free units to which this form can be traced back are various: the name *vetë* 'person', the adverb *vetë* 'by itself, without help'. In addition to the former, the attributive intensifier *i vet*, *e veta* 'own(M), own(F)', the noun *vete* 'self; groin', and the adjective *vetëm* 'only' also have affinities with the $_{CC}$marker, although they are partly distinct from it in terms of form. In the sentences below, the reflexive construction (104), the adnominal intensifier (105), and the exclusive adverbial intensifier (106) are exemplified.

it was already in the classical era that phonological factors caused a common confusion: the contracted form *hautón* of the reflexive *heautón* was often confused with the intensifier *autón,* once the aspiration has lost its distinctive value. The reflexivization strategy of Modern Greek emerges in the Byzantine period: first the form *heaut-* – only inflected for case – is used; then the current paradigm of forms inflected for person, number and case established itself through a complex process of reanalysis.

96 The lexicographic sources consulted for Albanian are sqB1, sqB2 and sqB3.

97 The base is derived from the verb *rrjedh* 'to flow, run' and the suffix *-je* which forms action nouns.

98 The form *vet-* is an allomorph of *vetë-,* which occurs in correspondence with bases starting with a vowel by deletion of the labile vowel *-ë.* The derivative is obtained from the base *asgjë* 'null', followed by the causative verbal suffix *-s(oj)*, which is then turned into an action noun by the suffix *-im*.

(104) (Buchholz and Fiedler 1987: 282; gloss: TDIR)
unë e lavdërova veten
I it praise REFL
'I praise myself.'

(105) (Buchholz and Fiedler 1987: 283; gloss: TDIR)
ajo vetë më tha
she INT to.me said
'She herself told me about it.'

(106) (Buchholz and Fiedler 1987: 283; gloss: TDIR)
plaga u mbyll vetë
wound MID heal int
'The wound healed by itself.'

Regardless of the different forms of the different free morphemes and lexemes identified as possible sources of the Albanian $_{CC}$marker, all of them have a cognate antecedent recognizable in the root *vet-*.[99] At any rate, the intensifiers with the invariable form *vetë* display the greatest formal correspondence with the $_{CC}$marker.

4.2.7 Uralic family

The Uralic languages included in the sample belong to the Finnic (*4.2.7.1*) and to the Ugric (*4.2.7.2*) branches of this family.

4.2.7.1 Finnish and Estonian

Let us start with Balto-Finnic languages: Finnish exemplified in (107).a and Estonian in (107).b.[100] These two genealogically very close languages present – unlike the Indo-European languages of the Celtic group, and even more so of the

[99] Orel (1998: 498) traces *vetë* back to Proto-Albanian **swai-ta*, derived from **swaja*, dating back to Indo-European **su̯oi̯os* 'own' (Old Slavic **svojь*, Old Prussian *swais*), linked in turn to **su̯e-*. Pokorny (1959: 882) traces *vetë* back to Indo-European **sue-ti-*.

[100] For Finnish and Estonian, the following lexicographic sources have been consulted: fiB1, eeB1 and eeB2.

Baltic languages group – etymologically related $_{CC}$markers.

(107) a. *itse·hallinto* 'self-administration [Selbstverwaltung]';[101] *itse·kylväyty-nyt* 'self-sown [selbstgesät]'[102] [fiB1]

b. *ise·majandamine* 'self-financing'; [103] *ise·toimiv* 'self-acting'[104] [eeB1]

The Finnish $_{CC}$marker *itse-* corresponds to the adnominal intensifier (108), to the adverbial one (109), and to the mark used in the reflexive construction (110).

(108) (TDIR)
presidentti *itse* *tervehti* *meitä*
president.NOM INT.NOM greeted.3SG us.PRTT
'The president himself welcomed us.'

(109) (TDIR)
korjasin *autoni* *itse*
repaired.1SG car.1SG.POSS INT.NOM
'I repaired my car myself.'

(110) (TDIR; with modifications)
hän *puolusti* *itse-ä-än*
he defend.PST.3SG INT-PRTT-POSS.3SG
'He defended himself.'

Finnish *itse*, as seen in the constructions in (108) and (109), inflects for case and agrees with the elements to which it refers, i.e. the elements under focus in the case of the adnominal and adverbial intensifier. In addition to case markers, *itse* also displays possessive affixes, agreeing with the participant with which the

[101] The base *hallinto* 'administration [Verwaltung]' derives from *hallita* 'to dominate (over) something [etwas beherrschen, über etwas herrschen]', with the suffix *-nto*/*-ntö* forming action nominals.

[102] Derived from *kylvää* 'sow [säen, aussäen, besäen]', followed by the reflexive suffix *-utu* + *a* / *-yty* + *ä* (Fromm 1982: 134), followed by the suffix of the active perfect participle *-nut*/*-nyt* (Fromm 1982: 118).

[103] From the verb *majandama* 'manage (an activity, a shop)', and the suffix forming action nouns *-mine*.

[104] From the verb *toimima* 'act, operate', and the suffix forming active present participles *-v*.

reflexive is coreferent in the case of the reflexive construction.

Estonian *ise* has the same range of functions as the Finnish *itse*, but as adnominal intensifier (111), adverbial exclusive intensifier (112) and as reflexive (113), it shows only case agreement, while it does not display possessive affixes.

(111) ([eeB1]; our gloss)
meie *ise*
1PL.EMPH INT.NOM
'We ourselves'

(112) ([eeB1]; our gloss)
me *ise* *teeme* *seda*
1PL INT.NOM do.1PL DEM.PRTT.SG
'We do it ourselves'

(113) ([eeB1]; our gloss)
me *peseme* *ennast*
1PL wash.1PL INT.PRTT
'We wash ourselves'

A last similarity unites Finnish and Estonian, and that is the presence of a secondary $_{CC}$marker, the attributive intensifier *oma,* whose source has meaning and functions similar to those displayed by *eigen*, and by cognate attributive intensifiers in other Germanic languages.

It should also be added that Estonian *ise* has a more complex inflection compared to Finnish *itse*, as can be gathered from its partitive form *ennast* in example (113). The presence of a richer paradigm has consequences in word-formation as well. In Estonian, the genitive is used as a unit in compounding and derivation along with the nominative. Compare, for example, the compounds *käsi-kiri* 'manuscript' (= hand.NOM-letter) and *käe-kiri* 'handwriting' (= hand.GEN-letter) (Raun and Saareste 1965: 36). Concerning derivation, see, for example, *kiriku-line* 'church-goer', where *kiriku* is the genitive of *kirik* 'church'. In 5.2.1 below, we will see that in Estonian the forms of the paradigm of *ise* which are different from that of the genitive occur as $_{CC}$markers.

4.2.7.2 Hungarian

Hungarian – the only representative of the Ugric group of the Uralic family in Europe[105] – differs from the other Finno-Ugric languages considered here, since the $_{CC}$marker *ön-* (114) in Hungarian does not have any etymological link with the ones (*ise* and *itse*) found in Estonian and Finnish.

(114) *ön·védelem* 'self-defence'; *ön·tömjénezés* 'self-praise'[106] [huB1]

It must be noted that we have identified two formations with *maga-* that indicate this as a candidate $_{CC}$marker, i.e. *maga·tehetetlen* 'non self-sufficient, totally incapacitated', and *maga·mutogatás* 'exhibitionism'. However, this word-formation pattern does not appear to be productive, as to the best of our knowledge it occurs only in these two lexemes, while *ön-* occurs in a long series of formations.[107]

As such, the closest syntactically free correspondent of *ön-* is the personal pronoun *ön*/*Ön*, a pronoun of formal address.[108] On the other hand, if we consider the functions of adnominal intensifier (115), adverbial intensifier (116) and reflexive (117), the marker used instead is *maga*.

(115) (TDIR)

mag-a	*az*	*igazgató*	*fogad*	*minket*
INT-POSS.3SG.NOM	ART	director	welcome	us

'The director himself welcomes us.'

[105] Apart from Hungarian, the other extant representatives of the Ugric group of Uralic languages, i.e. Khanty and Mansi are spoken outside the limits of geographical Europe (see Csepregi 2023: 704 and Virtanen & Horváth 2023: 665).

[106] The action noun *védelem* derives from the verb *véd* 'to defend', with the suffix *-alom*/*-elem*. The noun *tömjénezés* derives from *tömjénez* 'to incense; to flatter', a denominal verb issued from *tömjén* 'incense'.

[107] The compound *maga·tehetetlen* literally means '(someone who) can (*-het*) not (*-tlen*) do (*te-(sz)*) on his own (*maga*)' (compare *tehetetlen* 'incapable'). The compound *maga·mutogatás* is an action noun derived from the verb *mutat* 'to exhibit' through the suffix *-gat*, which has an iterative-continuous or habitual value, and therefore literaly means 'exhibition of oneself'.

[108] On the functions of *ön* in diachrony see the brief discussion in 6.1.1 below.

(116) (TDIR)

a	*leveleimet*	*mindig*	*(saját)*	*maga-m*	*bontom*	*ki*
ART	my.letters	always	(own)	INT-POSS.1SG.NOM	I.open	PRT

'I always open my letters myself.'

(117) (TDIR)

a. *meg-védte* *magá-t*
PRT-he.defended INT-POSS.3SG.ACC
'He deFENded himself.'

b. *saját* *magá-t* *védte* *meg*
own INT-POSS.3SG.ACC he.defended PRT
'He defended himSELF.'

In examples (116) and (117), *maga* can be used alone as the adverbial intensifier or to express reflexivity. However, the constructions with *maga* alone alternate with constructions in which the attributive intensifier *saját* 'own' is also present – see (116) and the contrast between (117).a and b. In addition, other constructions are also available, for example, *maga* can occur along with the third-person masculine pronoun *ő,* or in a compounded form with the aforementioned pronoun *ön*: *önmaga*. Consequently, the picture that emerges in Hungarian is similar to that of Modern Greek: the source of the $_{CC}$marker is a pronoun, while the reflexive marker (also corresponding to the intensifier in the case of Hungarian) may contain the same element but in a bound form.

4.2.8 Turkic family: Turkish

Turkish displays a picture similar to that of Hungarian and Greek. Although the $_{CC}$marker *öz-* is native, the elements that fulfill the function of intensifier or of reflexive have no formal correspondence with it. However, a lexeme belonging to the lexical pool of possible sources of intensifiers formally corresponds to it.[109]

In Hungarian and Greek, the syntactically free correspondents of $_{CC}$markers are part of the inventory of pronominal forms, they therefore appear to be at a more

[109] For Turkish, the lexicographic sources consulted have been tkB1, tkB2 and tkB3.

advanced stage of grammaticalization in relation to intensifiers. In Turkish, however, the source of the $_{CC}$marker is (synchronically) at the opposite end of the intensifiers-reflexive continuum, and less grammaticalized in relation to intensifiers, as *öz* can be a noun meaning 'the self, the essence', or an adjective meaning 'pure'.

(118) *öz·denetim* 'self-control [Selbstkontrolle]';[110] *öz·yönetim* 'self-management [Selbstverwaltung]';[111] *öz·indukleme* 'self-induction [Selbstinduktion]'[112] [tkB3]

Wilkens (2016: 3369) reports on the debate concerning the existence of prefixes in Turkish, and lists among others *öz-* as a problematic element whose status as prefix or as compounding element is a matter of dispute. He nonetheless advocates for treating the formations with *öz* and other elements like *baş* 'first' (e.g. *baş·parmak* 'thumb' lit. 'first-finger') as cases of compounding because they «are formed by juxtaposition with a following noun» (Wilkens 2016: 3369). We might add that *öz* can participate in derivatives such as *öz·deş* 'identical', formed with the suffix *-Taş*, which derives adjectives designating the characterization denoted by the base.

4.2.9 Afroasiatic family: Maltese

As anticipated in 3.1.8, although Maltese belongs to the group of Semitic languages and as such presents an introflexive, "root-and-pattern" morphology, the $_{CC}$marker active in word-formation in Maltese belongs to the concatenative layer of Italo-Romance origin which is integrated in the lexicon of this language. The marker in question, *awto-*, which can be easily identified in the examples in (119), is akin to the $_{CC}$markers attested in the Romance languages and in Modern Greek.

(119) *awto·difiża* 'self-defence'; *awto·regolazzjoni* 'self-regulation' [mtB1], [mtB2]

[110] From *denetim* 'control', linked to the verb *denetlemek* 'to control'.

[111] From the action name *yönetim* 'leadership, management', derived from the verb *yönetmek* 'to manage, to lead'.

[112] From *indükleme* 'induction', linked to the verb *indüklemek* 'to induce'.

The constituents on the right side of the formations in (119) also belong to the Italo-Romance stratum of Maltese. One could therefore ask why *awto-* is considered here as a productive word-formation pattern. There were two facts that lead to the conclusion that *awto-*, despite its stratal limitations, is indeed productive. First of all, although Maltese *awto-* is generally less productive than the related processes we can find in Romance languages, it is nonetheless evident that it does not occur exclusively in formations modeled on Italian. It can be argued that English is actually the model for the examples in (120): the choice of the form of the verbal base is calqued on English, while other morphological elements, i.e. the $_{CC}$marker and the suffixes forming action nouns, are clearly derived from the Italo-Romance layer. Concerning the latter, a suffix like *-(j)ar* – which is actually derived from the ending of the Italian infinitive *-are* – is now a productive suffix which forms deverbal action nouns, and it is fully integrated into the morphological system of Maltese (120).b.

(120) (source: EUR-Lex)[113]

a. *awto·evalwazzjoni* *self-evaluation*
b. *awto·assess·jar* *self-assess·ment*

The fact that a formation with *self-* in English is (as is evident from the example) transferred into Maltese using another element, i.e. *awto-*, shows on the one hand the ability of speakers to interpret and deconstruct the English model, but on the other, it shows that they are also able to interpret Italian models that display the prefix *auto-,* and then to productively use the $_{CC}$marker.

The second reason why we tend to consider Maltese *awto-* as a productive pattern is related to the fact that it is used, at least in a limited number of formations, with bases belonging to the Semitic stratum: see, for example, *awto-tmexx-ija* 'self-governance', *(materjali ta') awto·tiswija* 'self-repairing (materials), lett.

[113] The EUR-Lex is a multilingual portal that provides access to the laws and general legislation of the European Union. As such, it contains parallel texts translated into all the official languages of the States of the Union. The limit of this resource lies precisely in the fact that these are translations, which often draw heavily on the texts in the official languages of the EU (French and English). As for Maltese, the quality of the data collected through the EUR-Lex has been checked on sites of the .mt domain. The EUR-Lex (in this case the Maltese version) can be consulted at the website https://eur-lex.europa.eu/homepage.html?locale=mt (last accessed 2/7/2021).

(materials of) self-repair' (EUR-Lex).[114]

As for Romance languages, since *awto-* is a borrowed element in Maltese, there is no comparability with any syntactically free unit, either belonging or not belonging to the intensifiers-reflexive continuum.

4.2.10 Kartvelian family: Georgian

In Georgian, we have different $_{CC}$markers identifiable from the examples in (122).

(121) a. *tvit·mk'vleloba* 'suicide [Selbstmord]' [geB1][115]
b. *tavis·martleba* 'self-justification' [geB2][116]
c. *tav·dadeba* 'self-sacrifice [Selbstaufopferung]' [geB1];[117]

The possible syntactically free sources of these markers can be identified both in the domain of intensifiers and reflexives. We can see in (122) and (123) that the adnominal intensifier and the adverbial exclusive intensifier can assume one of the following forms: *tviton*, *tvit* or *tavad*. In (124), we have an example of the reflexive construction with (POSS) *tavi* 'PRO-self'.

(122) (Hewitt 1995: 85; gloss: TDIR)

me	*beč'ed-i*	*miveci*	*t(v)iton/tvit/tavad*	*dedopal-s*
I.ERG	ring.NOM	gave.it.to.her	INT	queen.DAT

'I gave the ring to the queen herself.'

(123) (Hewitt 1995: 84; gloss: TDIR)

tviton/tvit/tavad	*c'avida*
INT	s/he.went

'S/he went him/herself.'

[114] In addition to the examples cited, we can add *awto·tagħlim* 'self-teaching study', judged by a native speaker of Maltese as a possible, but not attested word (Sandro Caruana, p.c.).

[115] From the verb *mok'lavs* 'will kill sb/sth; will make sb very depressed' [geB2].

[116] See *martldeba* 'is justified, vindicated (by evidence, etc)', and *martleba* 'justifying, vindicating; fulfillment' [geB2].

[117] From *dadeba* 'putting down', an action noun derived from *debs* 'places, puts something down' [geB2].

(124) (Hewitt 1995: 563; gloss: TDIR)

(šen-s)	*tav-s*	*mo-i-k'l-av?*
(your-AGR)	head-DAT	PREV-SV-kill-FUT

'Will you kill yourself?'

Following Harris (1981),

> «Reflexivization is expressed in Georgian with the pronominal element *tav-*. This root also functions as an independent noun meaning 'head' or 'source'; it will be glossed here as 'self' when it functions as a reflexive. The forms *tavis-* (singular) and *taviant-* (plural) serve as possessive reflexives. The first person singular reflexive is *čemi tavi* 'my self', the second *šeni tavi* 'your self', and the third person reflexive is *tavisi tavi* 'self's self', that is, 'himself, herself, itself', without distinction of gender». (Harris 1981: 23)

The same noun *tavi* 'head' is involved in the formation of intensifiers, as Hewitt (1995: 84) notes. The forms *tv-it* and *tav-ad* are syncopated forms of the instrumental and of the so-called adverbial form of *tavi,* respectively – compare *gutan-i* 'plough[SG]-NOM' with *gutn-it(a)* 'plough[SG]-INSTR' and *gutn-ad(a)* 'plough[SG]-ADV' (Hewitt 1995: 35).

We can suppose that the more grammaticalized functions of the intensifier and the reflexive can still be recognized in synchrony as related to the noun *tavi*. On the other hand, in compounds such as *tav·daburuli* 'with head covered/veiled' and *tavis·at'k'iveba* 'headache' [geB2], the base used corresponds, respectively, to the root *tavi* 'head' and to its genitive form, a word-form which is not syncopated. In formations with *tvit-*, which appears to be the most productive $_{CC}$marker in Georgian with over 200 $_{CC}$formations in geB2, there is no competition between the interpretation 'head' and that of $_{CC}$marker, while the link to the intensifier is stronger since the $_{CC}$marker and the intensifier are phonologically identical.

4.2.11 Isolated languages: Basque

Basque displays a situation similar to that of Georgian, with a $_{CC}$marker that can be connected both to the lexeme for 'head', and to the marks used for the intensifier and the reflexive. The $_{CC}$marker of Basque is *buru-*, as can be observed in the

examples in (125).[118]

(125) *buru·salapen* 'self-accusation [autoacusación]';[119] *buru·hornikuntzka* 'self-sufficiency [autoabastacimiento]'[120] [euB1]; *buru·hilketa* 'suicide [Selbstmord]' [121] [euB2]

Unlike in Georgian, the syntactically free element most similar to the $_{CC}$marker is the noun *buru* 'head', which is also directly involved in the reflexive, as can be seen from the example (126). The marker *bere*, which instead acts as adnominal intensifier (127), is indeed etymologically connected to *buru* 'head', but it has a rather wide range of functions, including that of the third person emphatic pronoun (similar to Modern Greek *aftó*).

(126) (Saltarelli 1988: 104; gloss: TDIR)

aita-k	*bere*	*buru-a*	*hil*	*du*
father-ERG.SG	his	head-ABS.SG	kill	3SG.ERG.3SG.ABS.PRS.AUX2

'Father has killed himself.'

(127) (TDIR)

aita	*santua-k*	*bera-k*	*gurekin*	*hitzegino*	*du*
father	holy-ERG.SG	INT-ERG	we.SOC	talk.FUT	3SG.ERG.AUX2.PRS

'The Pope himself will talk to us.'

The adverbial exclusive intensifier is instead obtained by means of another construction that uses intensive personal pronouns (128). Pronouns, either ordinary or intensive, are available in Basque only for first and second person, and instead of third person pronouns, demonstratives are used (Hualde and Ortiz de Urbina 2003: 151). However, Western varieties of Basque have also acquired third person pronouns, which are derived from the third person intensive pronouns *bera* (SG) and *berak* (PL). In some

[118] In addition to formations with *buru-*, the marker *auto-* also occurs in Basque. However, the formations containing the marker *auto-* appear to be simply phonologically integrated loanwords from Spanish: *autoerretratu* < Spanish *autorretrato* [euB1].

[119] The base is *salapen* 'complaint, accusation', derived from the verb *salatu* 'denounce, betray'.

[120] The base *hornikuntzka* 'supply, endowment' derives from the verb *horni(tu)* 'deliver, provide, supply'.

[121] The base *hilketa* 'killing, murder' derives from the verb *hil* 'die; kill'.

Eastern varieties, intensive pronouns, including the third person *bera(k)*, have acquired the meaning 'by X-self, alone' (129) (Hualde and Ortiz de Urbina 2003: 153).

(128) (Saltarelli 1988: 134; gloss: TDIR with modifications)

ni-k	*neu-k*	*ikus-i*	*nuen*
1SG-ERG.SG	1SG.INT-ERG	see-PFCT	1SG.ERG.3SG.ABS.AUX2.PST

'I saw it myself.'

(129) (Hualde and Ortiz de Urbina 2003: 153; our gloss)

Pello	*bera*	*bizi*	*da*
Pete	3SG.INTS	live	3SG.AUX1.NPST

'Pete lives alone'

4.3 Typology of $_{CC}$markers in relation to lexical sources

From the overview of European languages presented so far, some general trends can be observed that allow us to establish a typology of the word-formation patterns involving $_{CC}$markers. These trends concern:

- the possibility of tracing the $_{CC}$markers used in the various languages back to specific syntactically free correspondents (lexical or grammatical);
- the inclusion of these correspondents into major syntactic categories or in the intensifiers-reflexive continuum;
- the similarity between the meaning of the $_{CC}$markers with the contribution of the syntactically free correspondents in relevant constructions.

As shown above, most $_{CC}$markers used in European languages can be traced back to syntactically free elements. When this occurs, we can, as a first approximation, define the relevant formations with $_{CC}$markers as some kind of compound. We can build Table 12 by providing a first rough subdivision of the word-formation patterns observed in the languages of Europe. Table 12 lists only the primary $_{CC}$markers – that is, the patterns that are applied to the largest set of different formations – and reports of the existence of a corresponding syntactically free element ("+" if present, "–" if absent).

In the table, for Romance languages and for Breton, we also chose to show a potential correspondence – although it is only formal – with a syntactically free element by using a plus within brackets "(+)", so that their case could be distinguished from that of Maltese where no correspondence, not even formal, is recognizable.

Table 12. $_{cc}$Markers in European languages. Availability of a syntactically free correspondent

Language	$_{cc}$markers	
Italian	*auto-*	(+)
French	*auto-*	(+)
Portuguese	*auto-*	(+)
Rumanian	*auto-*	(+)
Spanish	*auto-*	(+)
German	*selbst-*	+
English	*self-*	+
Dutch	*zelf-*	+
Danish	*selv-*	+
Icelandic	*sjálf-*	+
Swedish	*själf-*	+
Croatian	*sam-(o)-*	+
Serbian	*sam-(o)-*	+
Czech	*sebe-*	+
	sam-(o)-	+
Polish	*sam-(o)-*	+
Russian	*sam-(o)-*	+
Irish Gaelic	*féin-*	+
Scottish Gaelic	*féin-*	+
Welsh	*hunan-*	+
Breton	*em-*	(+)
Lithuanian	*savi-*	+
Latvian	*paš-*	+
Modern Greek	*afto-*	+
Albanian	*vet(ë)-*	+
Finnish	*itse-*	+
Estonian	*ise-*, ecc.	+
Hungarian	*ön-*	+
Turkish	*öz-*	+
Maltese	*awto-*	–
Georgian	*tvit-*	+
	tavis-	+
	tav-	+
Basque	*buru-*	+

Table 13 lists the various points of the intensifier-reflexive continuum, supplementing them with points outside the continuum, which are nevertheless relevant for the $_{CC}$markers observed earlier in the languages of Europe. The relevant parts of the continuum are provided with a number from 1 (*lexemes from major syntactic categories*) to 6 (*personal pronouns*), where the extremes are actually external to the intensifiers-reflexive continuum (also 2, *lexical sources of intensifiers*), while the central part belongs to the intensifiers-reflexive continuum proper: 3, *intensifiers*; 4, *stressed/heavy reflexives*; 5, *unstressed/light reflexives*.

The lexical sources from which the grammaticalization cline proceeds are considered outside the continuum in order to underline the fact that they belong to major syntactic categories and, therefore, to stress their low degree of grammaticalization. From this point of view, various types of intensifiers can have different degrees of grammaticalization, depending on whether they are variable (with characteristics similar to adjectives or pronouns) or invariable. However, since they provide a rather grammatical contribution in the sentences in which they occur, it has been considered to separate them from the lexical sources from which they emerge. At the other end of the continuum, we have personal pronouns. They are also connected with the intensifier-reflexive continuum as they constitute a possible evolution from intensifiers. However, given that their functions diverge from the ones that characterize the central part of the continuum, and most of all from the functions of $_{CC}$markers, they have also been considered outside the continuum.

Table 13. The intensifiers-reflexive continuum and other domains relevant for $_{CC}$markers

outside the continuum	*lexemes from major syntactic categories*	1
	lexical sources of intensifiers (BODY, HEAD, PERSON, ecc.)	2
intensifiers-reflexive continuum	*intensifiers*	3
	stressed/heavy reflexives	4
	unstressed/light reflexives	5
outside the continuum	*personal pronouns*	6

By combining the classification of Table 12 with the numbered subparts of the intensifiers-reflexive continuum as defined in Table 13, we can obtain the more insightful classification of Table 14. When the syntactically free element which corresponds to a $_{CC}$marker belongs to one of major syntactic categories, the $_{CC}$marker in that language is formally closer to prototypical compounding.

Conversely, when it corresponds to an element on the intensifier-reflexive continuum, then the $_{CC}$marker in that language can be considered semantically closer to prototypical compounding.

For each $_{CC}$marker, Table 14 shows the type of extension on the continuum defined above, and lists any additional functions not explicitly specified in the continuum. In this table, the symbol "(+)" indicates that the formal correspondence between the $_{CC}$marker and the syntactically free elements characterized by these functions is partial and lower, in relation to those corresponding to the functions indicated by "+". The symbol "[+]" indicates that there is formal correspondence with a syntactically free element, and that there is only a partial or completely absent semantic correspondence. From the point of view of diachronic development and grammaticalization, however, syntactically free elements, such as Romance AUTO 'car', have an inverse directionality to the usual one, as they are lexemes which develop from an element with a grammatical function (even if it is arguable to consider this a case of degrammaticalization). For Breton, the situation is more complex since it can be disputed that *em* in the reflexive construction *en em* is actually a syntactically free element, given its fixed position before the verb and its origin as a verbal prefix (see above in 4.2.4).

TABLE 14. $_{CC}$Markers in European languages and the intensifiers-reflexive continuum

Language	$_{CC}$markers	1	2	3	4	5	6
Italian	*auto-*	[+]	–	–	–	–	–
French	*auto-*	[+]	–	–	–	–	–
Portuguese	*auto-*	[+]	–	–	–	–	–
Rumanian	*auto-*	[+]	–	–	–	–	–
Spanish	*auto-*	[+]	–	–	–	–	–
German	*selbst-*	–	[+]	+	–	–	–
English	*self-*	–	[+]	(+)	(+)	–	–
Dutch	*zelf-*	–	[+]	+	(+)	–	–
Danish	*selv-*	–	[+]	+	–	–	–
Icelandic	*sjálf-*	–	[+]	+	–	–	–
Swedish	*själf-*	–	[+]	+	–	–	–
Croatian	*sam-(o)-*	–	–	+	–	–	–
Serbian	*sam-(o)-*	–	–	+	–	–	–
Czech	*sebe-*	–	–	–	+	–	–
	sam-(o)-	–	–	+	–	–	–
Polish	*sam-(o)-*	–	–	+	–	–	–
Russian	*sam-(o)-*	–	–	+	–	–	–

Language	$_{CC}$markers	1	2	3	4	5	6
Irish Gaelic	*féin-*	–	–	+	+	–	–
Scottish Gaelic	*féin-*	–	–	+	+	–	–
Welsh	*hunan-*	–	–	+	+	–	–
Breton	*em-*	–	–	–	–	–	–
Lithuanian	*savi-*	–	–	–	+	–	–
Latvian	*paš-*	–	–	+	–	–	–
Modern Greek	*afto-*	–	–	–	(+)	–	[+]
Albanian	*vet(ë)-*	–	[+]	+	+	–	–
Finnish	*itse-*	–	–	+	+	–	–
Estonian	*ise-*, ecc.	–	–	+	+	–	–
Hungarian	*ön-*	–	–	–	(+)	–	+
Turkish	*öz-*	–	[+]	–	–	–	–
Maltese	*awto-*	–	–	–	–	–	–
Georgian	*tvit-*	–	–	+	–	–	–
	tavis-	–	[+]	–	(+)	–	–
	tav-	–	[+]	–	(+)	–	–
Basque	*buru-*	–	[+]	(+)	+	–	(+)

Considering the data in Table 14, we can observe some synchronic trends of $_{CC}$markers. The first concerns the fact that none of the European languages considered has a $_{CC}$marker whose source is a light reflexive. The absence of this possibility in the data can be explained by the fact that this type of markers – which are not stressed or even affixal – tend to have a much more rigid distribution in relation to the verbal phrase than, for example, intensifiers or tonic reflexives themselves. In this regard, at least in two European languages where intensifiers and reflexives differ – Czech and Lithuanian – it is precisely the tonic/heavy reflexive marker that is used as $_{CC}$marker. Most of the languages in the sample, however, select the intensifier (17 languages, 19 considering cases of partial correspondence); of these, 7 languages (plus two with partial correspondence) display the identity of intensifier and reflexive. Using syntactically free elements belonging to this central area of the intensifiers-reflexive continuum as $_{CC}$marker constitutes a case of remarkable correspondence between the function of a syntactically free element and that of the relevant $_{CC}$marker. This functional correspondence is taken as a semantic criterion defining $_{CC}$markers that display a higher degree of prototypicality as units of compounding.

For 10 languages of the sample there is a formal correspondence between the $_{CC}$marker and a syntactically free element that belongs either to a major syntac-

tic category – usually that of nouns – or to the lexical pool of possible sources of the intensifiers-reflexive continuum (see paragraph 4.1 above). For some of these languages – English, Basque – the formal correspondence with these elements external to the intensifiers-reflexive continuum is greater than that between the $_{CC}$markers and the intensifiers/reflexives. Concerning the case of English and Basque and the content of Table 14 in general, some clarifications are needed. On the one hand, Table 14 must be interpreted from a synchronic perspective. In this sense, even the fact that an element belongs to the various classes defined in Table 13 mirrors the synchronically observable situation in the various languages. However, if we take a diachronic perspective, and if we carefully consider the relative position of the individual elements considered in Table 14 on the diachronic axis, the cases of Basque and English (and more generally of Germanic languages) cannot be assimilated. Basque *buru* 'head' is actually the source of the intensifier *bere,* and this nominal use is diachronically prior to that of the intensifier. On the contrary, the English noun *(the) self* is the result of a substantivization process of an element already used as the intensifier, and the same holds true for German *(das) Selbst* (see Chapter 2, section 2.2.6) and for other Germanic languages. The development of the noun is thus subsequent to the use of the same lexeme as intensifier, and therefore, in this case a plus in square brackets '[+]' is used in correspondence to the relevant use as noun – position 2 of Table 13 – as in the case of Romance markers – for which *auto* 'car' is considered to hold position 1 of Table 13. These cases are nonetheless considered in the table because of the possible synchronic relevance for the formation of compounds with *self* and the cognate elements of other Germanic languages. As has already been observed several times (in Chapter 2 and above in 4.2.2), this formal correspondence is by itself no overall guarantee of a higher degree of prototypicality of a $_{CC}$marker as a unit of compounding, since the lexical traits of the nouns considered do not emerge in their use in compounds in which they convey contrastive coreference.[122] However, this correspondence does point towards a higher degree of prototypicality of a $_{CC}$marker as a unit of compounding from a **formal** point of view. This is even more evident in the case of Romance languages, where we observe the formal correspondence of $_{CC}$markers with a syntactically free element, which is nonethe-

[122] It should be noted that in Basque, *buru* can also occur in compounds with its lexical meaning 'head' – *buru-estalki* 'hood, cuff, lit. lid / blanket-head' [euB1].

less due to the accidental synchronic homonymy of AUTO- as a $_{CC}$marker and AUTO as a noun meaning 'car'. However, it should be noted that in the cases in which a correspondence between a $_{CC}$marker and a body-part lexeme occurs, usually a correspondence with the intensifier(-reflexive) is also present.[123] This shows how the purely formal and the semantic motivation of $_{CC}$markers are steadily competing: if an element belongs to a major syntactic category, the corresponding $_{CC}$marker will constitute a more prototypical compound unit from the formal point of view, but it will be less prototypical from the semantic point of view, and vice versa. This competition derives from the fact that the contribution of $_{CC}$markers is eminently grammatical: this is also demonstrated by the fact that, even if a marker belongs to a major syntactic category, it must have already developed at least one of the functions included in the intensifiers-reflexive continuum.

At the other end of the continuum, where intensifiers evolve into pronouns, we find Hungarian and Modern Greek. Also in this case, however, an at least partial correspondence with heavy reflexive markers (still) exists. For Romance languages, although there is a formal correspondence with syntactically free elements, this cannot be taken into consideration because it is not accompanied by any semantic relationship between the syntactically free element and the $_{CC}$marker. Furthermore, it does not correspond to any kind of grammaticalization path directed toward the intensifiers-reflexive continuum. Finally, concerning Maltese, *awto-* completely lacks formal correspondents. The case of Romance languages, Breton and Maltese constitutes probably the furthest point from prototypical compounding: the $_{CC}$markers found in these languages can be considered as part of derivation, and should be most probably regarded of as prefixes.

[123] The case of Turkish is an exception to this tendency, since the $_{CC}$marker *öz-* corresponds to a noun, *öz*, while the intensifier and reflexive markers are built using the element *kendi* (with the exception of a use of *öz* as attributive intensifier). See also Chapter 6, section 6.1.1.

4.4 Conclusion: Constructional schemas for contrastive coreference markers

Starting from the discussion of the sources of $_{CC}$markers in European languages, we have come to the survey included in Table 14. We can now interpret the table from the point of view of Construction Morphology, and identify a series of networks of lexical relations representing different patterns of motivation of the $_{CC}$markers identified.

The first type of $_{CC}$marker that we identify is constituted by all languages for which the source of the marker can be traced back to a noun that is also synchronically linked to elements lying on the intensifiers-reflexive continuum. The hierarchy for this type of $_{CC}$markers is drawn in Figure 13.

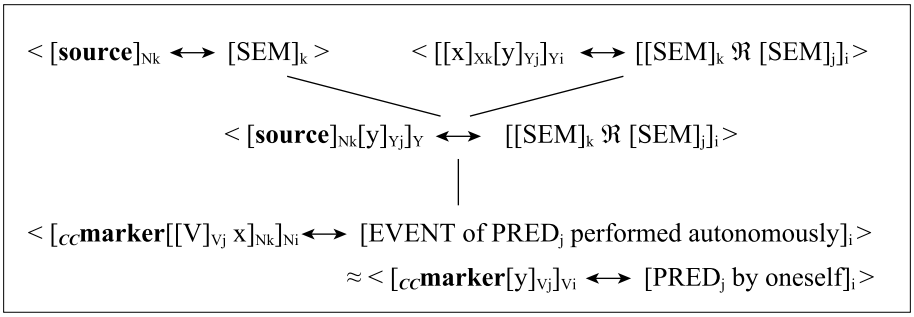


Figure 13. Compound $_{CC}$markers I

To this first type belong the $_{CC}$markers of all Germanic languages and of Albanian, Georgian *tavis-* and *tav-*, and Basque.[124] This first type of $_{CC}$markers is the closest to prototypical compounding, even if, as discussed above for German *selbst-* in Chapter 2, section 2.3.4, it also displays features that correspond to the disputed class of affixoids, i.e. a borderline case between compounding and derivation. We will call this type "Compound $_{CC}$markers I". We can also ascribe to this type the $_{CC}$marker of Turkish, with the slight difference that, unlike the $_{CC}$markers belonging to this type, Turkish *öz* cannot be linked to any function related to the intensifiers-reflexive continuum. However, given that *öz* in its syntactically free use belongs to the major word classes of nouns and adjectives, it is

[124] It must be noted that the Georgian markers *tavis-* and *tav-* correspond to the markers used in the reflexive construction rather than to the intensifier(s).

closer to prototypical compounding from a formal point of view. However, from the semantic point of view, the $_{CC}$marker of Turkish would be a good example of affixoid, because of the lack of any link with the intensifiers-reflexive continuum motivating the interpretation of the schema.

The second type of $_{CC}$markers is a step further toward derivation. In this case, as represented in Figure 14, we have a very similar situation as with "Compound $_{CC}$markers I", the difference being that it is not possible to find a syntactically free element which can be considered as a possible source of the $_{CC}$marker and is also a lexical source of intensifiers. In this case, the source of $_{CC}$markers is an element that belongs to a closed class and is a syntactically free grammatical marker.

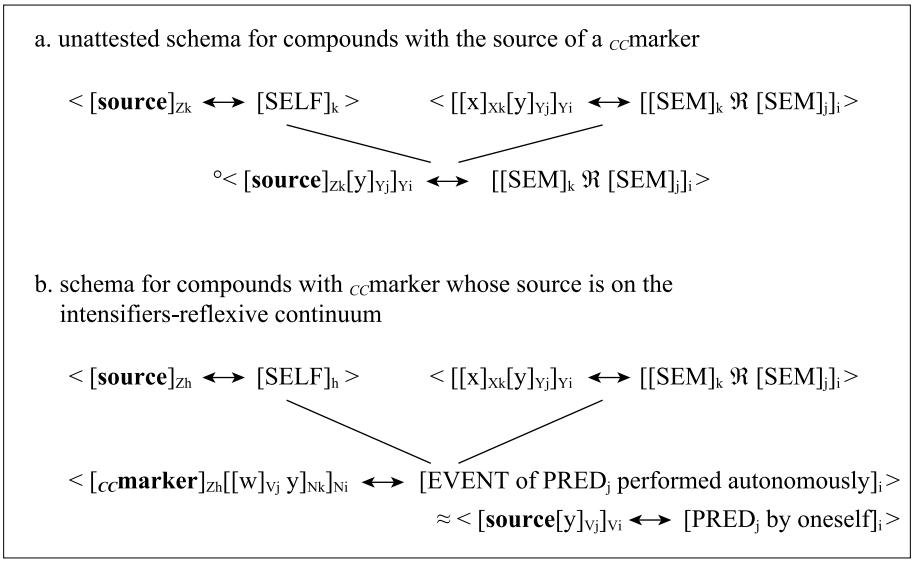


Figure 14. Compound $_{CC}$markers IIa

This is the case of Slavic languages (only SAM(O)- markers), Celtic languages (except Breton), Latvian, the Uralic languages of the Finnic group and Georgian *tvit-*.

We can call these markers "Compound $_{CC}$markers IIa". The letter "a" indicates that the markers of the following group, i.e. "Compound $_{CC}$markers IIb", only constitute a subtype of the previous group. In fact, the distinction between IIa and IIb lies in the fact that, while Type IIa finds part of its motivation in a paradigmatic relation with the syntactic construction of intensifiers, Type IIb is linked to the reflexive construction instead.

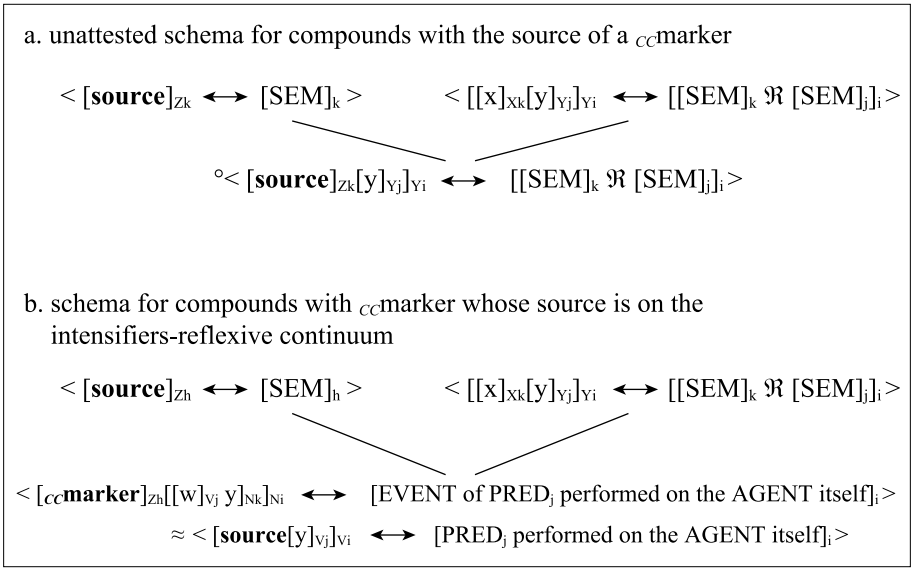


Figure 15. Compound $_{CC}$markers IIb

However, not all languages of the sample distinguish between the two functions – see e.g. Celtic and Uralic languages of Type IIa, but also English, Dutch, Georgian and Basque as far as Type I is concerned. Consequently, the $_{CC}$markers of Celtic (Breton excluded) and Finnic languages belong to both Type IIa and Type IIb, while SAM(O)- $_{CC}$markers of Slavic languages and Latvian *paš-* belong only to Type IIa. Czech *sebe-* and similar Slavic markers, along with Lithuanian *savi-* belong only to Type IIb.

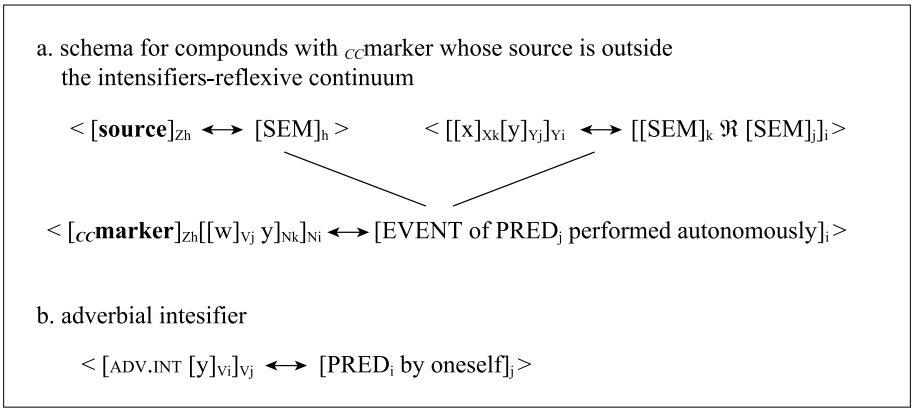


Figure 16. Compound $_{CC}$markers III

The third type, represented in Figure 16, is constituted by those markers which lack a syntactically free correspondent among the lexical sources of intensifiers, as well as a strong formal correspondence with any of the markers on the intensifiers-reflexive continuum. This is the case of Modern Greek and Hungarian. This type of markers in these languages have a partial correspondence with the reflexive, but display a stronger formal correspondence with a pronoun, i.e. a function that constitutes a possible output of intensifiers in their grammaticalization cline, but which lies outside the intensifiers-reflexive continuum.

Given the weak motivation network of Type III $_{CC}$markers, one could argue that they are almost affixal in nature. On the one hand, $_{CC}$markers of all types have a rather grammatical meaning and, even when referable to the domain of compounding, they display features that make them close at least to borderline cases such as that of affixoids. However, on the other hand, Type III $_{CC}$markers almost fail in linking to at least one of the functions present on the continuum of intensifiers and reflexives, if we exclude the partial formal link with reflexive markers. This makes them the type of compound $_{CC}$markers closest to derivation. This is very interesting in the case of Hungarian, given that in this language, apart from the marker of superlative and the large class of separable preverbs, no other prefix can be identified.

The last type that we identify is that of "Derivational $_{CC}$markers IV", represented by Romance languages, Breton and Maltese $_{CC}$markers.

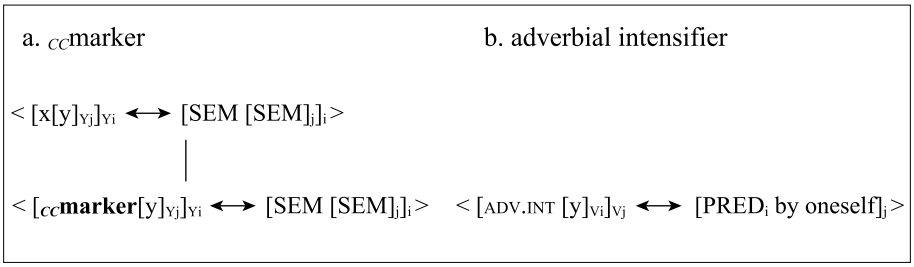


Figure 17. Derivational $_{CC}$markers IV

In this context, the Breton $_{CC}$marker is probably closer to derivation, given the affinity of Romance AUTO- with neoclassical combining forms, i.e. a type of compounding, though non-prototypical.

The last note concerns the complete absence of $_{CC}$markers whose source is a light reflexive marker. This is not unexpected, given the fact that these markers

often function more as valency changing devices (anticausative, passive, middle markers) than as reflexive markers establishing the coreference of clearly profiled participants of a transitive predicate.

In the next chapter, we will investigate the compatibility of $_{CC}$markers with different sets of base words in search for patterns of polysemy and of language-specific constraints.

5

COMPATIBILITY AND FUNCTIONS OF CONTRASTIVE COREFERENCE MARKERS

In this chapter, we will consider the bases with which $_{CC}$markers combine in various languages. In the previous chapter, we have examined $_{CC}$markers from the point of view of their correspondence with syntactically free elements in order to give as clear a picture as possible of the types of word-formation patterns they are involved in. We have concluded that these markers are placed on a continuum between compounding and derivation, but we have not been able to reduce them to one or the other unless we overlook at least some of their features, either on the formal or on the semantic level. In this sense, Construction Morphology has helped us to account for the nature of $_{CC}$markers by providing constructional schemas and inheritance links that can motivate their formal features. In this chapter, we will turn our attention from $_{CC}$markers to the bases of relevant word-formation patterns by examining the compatibility of $_{CC}$markers with different classes of words that constitute the bases of the resulting formations. We will try to induce some generalizations relating to these formations, based on the observation of this compatibility, by verifying whether the reciprocal position of various markers on the compounding-derivation continuum constitutes a predictive criterion with respect to the observed behaviours. We will also evaluate other features that may constitute predictive criteria.

5.1 Conceptual space of parts of speech

This section illustrates the conceptual space of parts-of-speech (henceforth POS), proposed by Croft (1991, 2000, 2001), and argues in favor of its use in the study of word-formation to complement the use of syntactic categories with a more fine-grained and cross-linguistically valid classification of word-formation units. We will provide the reader with examples of how word-formation rules and cases of compounding have restrictions deriving from the semantic class to which different bases belong rather than from their syntactic category. On the other hand, we will show how in compounding elements otherwise belonging to a syntactic category can convey meanings typical of elements belonging to other categories.

As mentioned earlier in Chapter 2, in word-formation rules forming new words (or lexemes) are characterized by a domain that is the pool of bases to which a specific rule applies. Constructional schemas, even if constituting a different the-

oretical object, display similar combinatorial limitations that involve the definition of sets of words that enter a specific schema. When we proposed a definition of compound in Chapter 2, we noted that there are some classes of bases that correspond to the syntactic categories to which relevant words/lexemes correspond. Major syntactic categories, also intended as classes of open elements – that is, not definable by their extension – are the categories of nouns, adjectives and verbs. In addition to these categories, the class of prepositions can also productively enter word-formation schemas, both in compounding (Italian *lungofiume*) and in derivation (see German *vorig* 'past; anterior' da *vor* 'before'), however, it will not be further treated in this section.

The definition of these syntactic categories is not primarily a product of morphological theory in word-formation, rather, it is inherited from syntactic theory. It is essentially based on morphosyntactic criteria, i.e. on the morphosyntactic exponents that the elements of each category display, and on distributional criteria, i.e. on the possibility to occur in certain syntactic constructions. However, on the one hand, the morphosyntactic criteria are not satisfactory in defining, for example, the class of nouns, whose elements can be uninflected even in an inflectional language, while they are simply inapplicable to isolating languages. On the other hand, distributional analysis can lead to a proliferation of POSs, and the more detailed the criteria on which it is based are, the less they are able to provide categories with clear boundaries. Furthermore, none of these groups of criteria, which are limited to a formal level, help to define cross-linguistically valid features on which to base an effective theory of POSs.

Croft's proposal for a theory of POSs from the cross-linguistic perspective is based on the idea that this theory must be based on typological principles that are motivated independently from language-specific considerations. On the one hand, the model recognizes a series of functional-pragmatic functions: REFERENCE, MODIFICATION and PREDICATION. The conceptualization provided by these functions is defined as follows:

(130) Croft (1991: 108)

REFERENCE creates an autonomous entity, and makes it into a kind or an individual of the kind (with its attendant stereotype and connotations);

MODIFICATION makes the entity into a stable but one-dimensional description

(property) of some other entity;

PREDICATION sequentially scans an entity, making it into a transitory (and changing?) state of affairs, involving at least one participant.

The traditional tripartite subdivision of parts of speech into the classes of nouns, adjectives and verbs is based on this distinction. Through the three pragmatic functions, it is possible to identify the main parts of speech, while overlooking more minute categories that are compatible with specific distributional criteria. Furthermore, through this functional criterion, it is possible to obtain the tripartition without defining specific criteria for individual languages.

On the other hand, Croft's proposal involves a purely semantic dimension that corresponds to the semantic classes of words of OBJECTS, PROPERTIES and ACTIONS. These classes of elements are defined by a set of semantic properties, which characterize prototypical members of the major syntactic categories. Semantic properties that characterize the three semantic classes of words are summarized in Table 15.

TABLE 15. Properties of semantic classes of words (Croft 1991)

	Objects	**Properties**	**Actions**
Valency	0	1	≥1
Stativity	state	state	process
Persistence	persistent	persistent	transitory
Gradability	nongradable	gradable	nongradable

The semantic properties in Table 15 emerge by taking into account the cross-linguistic effects of markedness: it has been observed that the elements belonging to the syntactic categories or nouns, adjectives and verbs, which are characterized by the semantic properties of OBJECTS, PROPERTIES and ACTIONS respectively, cross-linguistically display a number of morphemes which is not greater than the elements of the same classes that do not share those semantic properties.

The intersection of the functional-pragmatic functions listed in (130), and of the semantic classes of words characterized by the properties shown in Table 15, predicts the nine classes shown in Table 16. In Table 16, we observe three unmarked combinations of unmarked nouns, unmarked adjectives and unmarked verbs, and a series of other possible marked combinations that structure the conceptual space of parts of speech.

Table 16. Classes of parts of speech (Croft 1991)

	REFERENCE	MODIFICATION	PREDICATION
OBJECTS	UNMARKED NOUNS	genitive, adjectivalizations, PPs on nouns	predicate nominals, copulas
PROPERTIES	deadjectival nouns	UNMARKED ADJECTIVES	predicate adjectives, copulas
ACTIONS	action nominals, complements, infinitives, gerunds	participles, relative clauses	UNMARKED VERBS

The choice of an approach of this type is due to the consideration that the perspective generally held in word-formation has some operational and theoretical limitations. In the first place, it is not possible to apply syntactic tests to attribute a lexeme or a unit of compounding – used as a modifier – to a certain syntactic category. Just as in 4.2.7.1 it was not possible to decide whether a $_{CC}$marker such as Finnish *itse-* corresponds to the intensifier or the reflexive, it is similarly not possible to decide the syntactic category of compound modifiers that correspond to lexemes which can be linked to more than one syntactic category. Let us consider, for example, a compound like the German *glaubwürdig*: what is the syntactic category of the element *glaub-*? This element can easily be connected to the noun *Glaube* 'faith', or to the verb *glauben* 'to believe', thus allowing two paraphrases: 'worthy of faith' or 'worthy of being believed'. It can be objected that *Glaube* and *glauben* are related lexemes, and that it is indeed acceptable to consider the noun as a derivative by conversion from the verb, since it has the semantics of an action nominal – 'faith' = 'the fact of believing'. But this is precisely what limits a definition of the units of compounding by syntactic criteria: what is crucial in the case of *glaubwürdig* is not the syntactic category of the element *glaub-* – which is in fact undecidable[125] – but the semantic content of both *Glaube* and *glauben*, i.e. the fact that both lexemes, regardless of the syntactic category to which they belong, express an action.

For modifiers, the difficulty of applying tests to establish the syntactic category to which they belong is clear, although in a language such as Italian – which has left-headed compounds and compounding elements provided with inflectional exponents – it is often possible to observe some morphosyntactic features that

[125] The undecidability can be partially avoided by relying on indirect procedures: *seh·würdig*, where *seh-* can only be a verbal basis. This parallel is only partially decisive, because there are opposite cases, in which a modifier has action semantics, but it clearly corresponds to a noun: *kritik·würdig* is obtained by the noun *Kritik*, while **kritisier·würdig* is unattested.

depend on the syntactic category of those lexemes: for example, in *raccolta rifiuti* 'waste collection', the plural of the noun *rifiuto* occurs.

The advantage of a perspective in which semantic properties of lexemes are privileged over their membership in syntactic categories is less evident if one considers the elements that function as the head of compounds. As far as heads are concerned, it is always possible to apply syntactic tests on the entire compound, just as they can be applied to the corresponding lexemes in their free syntactic use, provided that the compound is endocentric, because endocentric compounds inherit the morphosyntactic traits of their heads.

As for the possible restrictions on the use of a certain element or class of elements in a compound, they can be understood as linked to the semantic class of the relevant elements. If we reinterpret the classes of compounds proposed in Bisetto and Scalise (2005) and Scalise and Bisetto (2009) – and the examples provided there – from the point of view of the conceptual space of POSs, it is possible to identify constraints or general tendencies in the formation of attributive, coordinative or subordinative compounds that do not depend on the syntactic categories of the bases, but are rather due to their membership within a specific semantic class. For example, if attributive compounds are considered from this perspective, they display the kind of constraints that is linked both to the fact that the modifier is an adjective (and most often indicates a PROPERTY) – e.g. *blackbird* – and to the fact that its head is a lexeme that is characterized by the trait [+ concrete] – making it an OBJECT in Croft's terms – to which it is therefore possible to attribute the property of being black. As for subordinate compounds (in particular the "verbal nexus" subtype), the head, generally a deverbal basis – and most often an ACTION in Croft's terms – will have an OBJECT as modifier – think of *can opener*.[126] If we focus on adjectival coordinative compounds, the constraint that the coordinates express a PROPERTY – *giallo-verde* 'green-yellow' – holds in the case of the coordination of relational adjectives as well. The coordination of relational adjectives – i.e. OBJECTS that have the propositional function of MODIFIERS – appears acceptable only in cases where the relational adjectives have already developed an interpretation as adjectives indicating a PROPERTY (*?l'auto presidenzial-senatoriale* 'the car of the president-senator', *un atteggiamento presidenzial-senatoriale* 'a presidential-senatorial attitude').

[126] These are obviously trends that characterize the most typical compounds of the various classes: the head of a subordinate compound is not necessarily a deverbal element expressing an ACTION; see *glaub·würdig* above, which also shows how the fact that the modifier is an OBJECT is not a restrictive constraint.

Obviously, in order to obtain more valid generalizations, the granularity of the semantic analysis of compounds should be further increased beyond the simple adoption of this perspective on the POSs that combines pragmatic and semantic criteria. Furthermore, it should be noted that the conceptual space of POSs has been conceived mainly for the cross-linguistic study of syntactic constructions, rather than of word-formation phenomena and the data that we have considered in the previous chapter and will deal with in this chapter – which are mostly taken from lexicographic sources – generally lack information on their uses in different syntactic and pragmatic contexts. For this reason, the dimension of the pragmatic function will inevitably be reduced to what we are able to obtain from dictionaries, which means that it will in the end correspond to the syntactic category as understood in traditional senses. Adopting the two-dimensional conception of parts of speech will at any rate be a useful starting point for obtaining a clearer and more reliable picture of the relationships between elements involved in word-formation. The data that will be presented in the next paragraphs will further clarify the usefulness of this perspective, especially in the case of a phenomenon such as that of the bases modified by a $_{CC}$marker, which, as we have already shown above, thrives at the boundary of derivation and compounding.

5.2 Interaction of contrastive coreference markers and bases

As argued in the previous paragraph, in order to produce appropriate generalizations with respect to the restrictions or, more generally, to the tendencies concerning base selection in complex formations built with a $_{CC}$marker, the level of their propositional function – i.e. the membership of a base within a specific syntactic category – and that of the semantic class to which these bases belong must be kept distinct.

In the next paragraphs, we will first consider the case of bases belonging on the one hand to the different major syntactic categories of nouns (REFERENTIAL function), adjectives (MODIFICATION) and verbs (PREDICATION), and on the other hand, to the semantic class of ACTIONS. We will then observe the compatibility of $_{CC}$markers with bases belonging to the semantic classes of PROPERTIES and OBJECTS.

It should be noted that in this paragraph we will not consider formations with AUTO- (and related phonological/orthographic outcomes, such as Turkish *oto-* and Russian *avto-*) in all those languages (including German) where this element is only part of neoclassical formations, and does not productively link to new bases

of the native lexical stock. A similar criterion will also be applied for the exclusion of all the neoclassical formations of Romance languages from this survey.

5.2.1 Syntactic categories of the base words

In Chapter 2, we have observed that in Italian and German the bases of the formations with *auto-* and *selbst-* mostly belong to the semantic class of ACTIONS. Furthermore, it was noted that deverbal action nominals were the most heavily represented pool of bases in the types of both languages. The two languages, however, differed concerning the other syntactic categories of adjectives and verbs. Angster (2012) shows the presence of a large number of deverbal adjective formations of various types – eventive adjectives, deverbal relationship adjectives, result adjectives – in both languages, but only Italian displays a substantial number of verbs in predicative function prefixed with *auto-*.

In European languages we observe a similar trend. All the languages considered show the presence of deverbal action nouns and deverbal adjectives as the basis of formations with $_{CC}$markers. Only for some languages, however, we attest the possibility to have verbs in predicative function displaying a $_{CC}$marker.

As already mentioned, in all languages of the sample, the $_{CC}$markers that combine to bases expressing an ACTION combine at least to bases that have referential function, i.e. those belonging to the syntactic category of nouns. This happens for both languages that, such as Italian and German, display a substantial number of patterns and formations involving $_{CC}$markers, and languages that instead display a limited number of formations, such as Turkish and Maltese.

In Figure 18, we show how Turkish and Maltese in fact display formations attributable exclusively to the area of the conceptual space of POSs at the crossing of REFERENCE × ACTION: e.g. Turkish *öz·indukleme* ‘self-induction’, Maltese *awto-kontraddizzjoni* ‘self-contradiction’. Most of the languages of the sample, however, have $_{CC}$markers that can combine to bases at the crossing of REFERENCE × MODIFICATION. Figure 18 shows German *selbst-* (*selbst·klebend* ‘self-gluing’, *selbst·ernannt* ‘self-named’), Finnish *itse-* (*itse·kylväytynyt* ‘sown/planted by itself’, *itse·valaiseva* ‘that lights up by itself’),[127] and Hungarian *ön-* (*ön·működő* ‘that

[127] «*itsevalaiseva* ‘selbstleuchtend’» [fiB1]. From the verb *valaist* ‘leuchten; beleuchten’, plus the present participial suffix *-va*, *-vä*.

works by itself',[128] *ön·szabályozó* 'self-regulating').[129]

Finally, it should be noted that the $_{CC}$markers of some languages of the sample – in particular of Romance languages, Modern Greek, Albanian – also connect to bases consisting of verbs in predicative use. We will return to this case below.

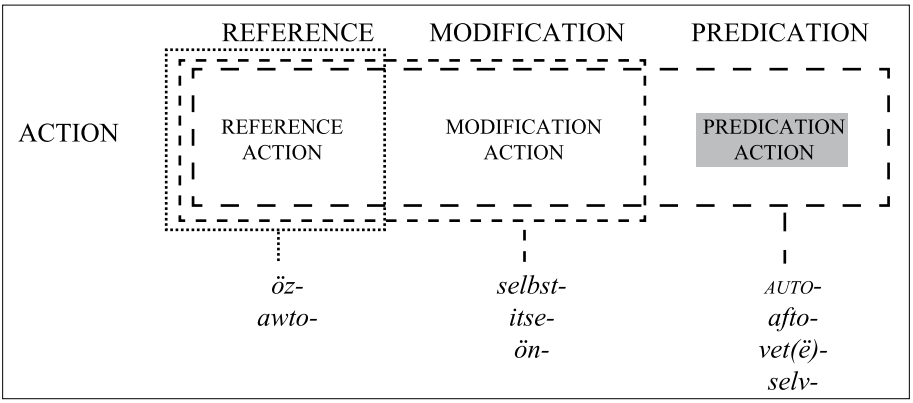


FIGURE 18. Semantic map (1): Compatibility of $_{CC}$markers with bases in the class of ACTIONS

A first approximation of this tendency is put forward in (131).

(131) **Tendency I**
If a language has a $_{CC}$marker, then this marker will occur at least with bases characterized by the features REFERENCE × ACTION.

This tendency does not have a high degree of generality, because it excludes those markers that apply exclusively to bases outside the semantic class of ACTIONS (see 5.2.3 below). Some markers of Slavic languages exemplify this case, such as Polish *własn-(o-)*, Croatian *svoje-* and *vlasn-(o-)*, the Czech markers *své-* and *vlastn-(o-)* and the Russian markers *svoe-* and *sobstvenn-(o-)*, occuring with bases placed in the class of OBJECTS or with states, which are actually a class that can be subsumed under PROPERTIES rather than under ACTIONS. Because of these cases, in order to propose a more general tendency, (132) can be formulated so that it accounts for at least languages with a single $_{CC}$marker:

[128] From the verb *működik* 'to function', with the suffix *-ő* that forms adjectives from verbs that can function as agent/instrument nouns as well.

[129] From the verb *szabályoz* 'to regulate', with the suffix *-ó* (see note 4).

(132) **Generalization I.a**

If a language has only one $_{CC}$marker, then this marker will occur at least with bases characterized by the features REFERENCE × ACTION.

The applicability of the generalization in (132) can then be widened to the languages in the sample that display more than one $_{CC}$marker if formulated as in (133).

(133) **Generalization I.b**

If a language has more than one $_{CC}$marker, the most productive one will occur with bases characterized by the features REFERENCE × ACTION.

The corollary in (134) can be then inferred from the second version of **Generalization I**:

(134) **Corollary to Generalization I.b**

If a language has more than one $_{CC}$marker, then, if one of them occurs exclusively with bases outside the semantic class of ACTIONS, it will not be the most productive $_{CC}$marker.

We will consider other implications of the corollary when discussing the more general tendencies related to the semantic classes of words. However, it is important to take this tendency into account already at this point to show how the possibility of having different $_{CC}$markers can correspond to differences in productivity and combinatorial possibilities.

In general, we have not observed any functional specialization between different markers in relation to bases expressing actions. In particular, if there are several $_{CC}$markers that can connect to bases of this type, in most of the languages of the sample we do not observe functional specialization corresponding to the functions of *auto-* and *selbst-* proposed in the literature and exposed in Chapter 1, section 1.2. This leads to the tendency described in (135).

(135) **Tendency II**

If a language has more than one $_{CC}$marker, then the different markers do not display any functional specialization across subgroups of bases belonging to the class of ACTIONS.

This tendency obviously excludes all languages with a single $_{CC}$marker (Romance languages, Breton, Latvian, Greek, Albanian, Hungarian, Maltese, Turkish, Basque),[130] but it involves most Germanic and Slavic languages. Exceptions to this tendency are English and Icelandic, among Germanic languages, Czech among Slavic languages, and Estonian among Finno-Ugric languages.

Competing $_{CC}$markers in different languages display different types of functional specialization. In some languages, functional specialization appears to be linked at least at first sight to the syntactic category to which bases belong. For example, in Icelandic the form *sjálfs-* (homonym of the genitive form of the intensifier) occurs preferably with nouns – about 90 nouns against 3 adjectives – while the form *sjálf-* (homonym of the nominative) occurs with a clear, but less overwhelming prevalence with adjectives, compared to nouns – about 51 against 29.[131] This state of affairs is represented below in Figure 19.

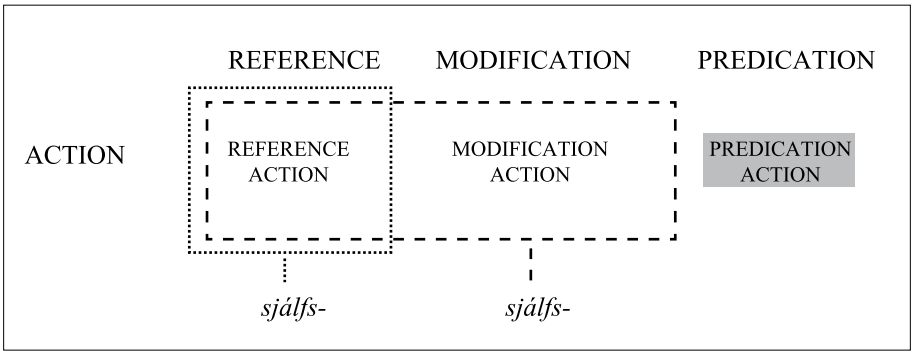


Figure 19. Semantic Map (2): functional specialization of Icelandic $_{CC}$marker sjálf(s)-

However, if we consider that in connection with adjectives an agent-oriented interpretation (the adverbial function in König 2011) is particularly common – for example *sjálf·losandi* 'self-detaching' – this specialization, which is at first sight linked to the syntactic category of the base, might depend on the interpretation on the functional level.

The case of Estonian is also connected to a distinction linked to the syntactic category. In Estonian, the main $_{CC}$marker is always homonym of the reflexive intensifier

[130] It should be reminded that in this work we do not consider the possible functional specialization between $_{CC}$markers and syntactic constructions – see, for example, the existence of *sicuro di sé*, but not of **autosicuro* in Italian.

[131] The quantitative data on Icelandic are taken from Jónsson (2005), a dictionary that lists sets of words characterized by the same morphological formant.

ise, but not always to its nominative form (*ise·oppimine* 'self-education', *ise·süttimine* 'spontaneous combustion'). Concerning one of the competing forms of the $_{CC}$marker – i.e. the genitive of *ise*, *enese* (*enese·süüdistus* 'self-accusation', *enese·mõrv* 'suicide') – its presence can be explained simply by observing that in Estonian the genitive form of nouns is frequently selected as the basic form for compounds and even for derivatives (Kerge 2016). Another form – *ennast*, homonym of the partitive of *ise* – shows instead a tendency towards the selection of adjectival bases: see, for example, *ennast·hellitav* 'self-indulgent' vs. *enese·hellitus* 'self-indulgence'; *enese·hävitav* 'suicidal (adj.)' vs. *enese·hävitamine* 'self-destruction'). This tendency is far from absolute – there are adjectives in which *enese-* occurs and nouns with *ennast-* – but it shows how Estonian exploits different forms of the intensifier-reflexive *ise* paradigm, which one might be tempted to interpret in a syntactic way. For example, the use of the partitive with adjective bases that are frequently eventive could correspond to the function of the partitive in Estonian of conveying the imperfectivity of an action with the use of a partitive object.[132] Other examples of semi-syntactic uses of the intensifier *ise* in compounds are *enesesse·sulgumine* 'introversion', where we have the illative form of *ise*; or *eneste·vaheline* with the plural genitive and a meaning corresponding to that of *oma·vaheline* 'with one's own means' – with *oma*, the attributive intensifier. This variety of different forms and the tendency to use them almost syntactically makes the Estonian formations with $_{CC}$markers closer to syntax and far from prototypical compounding. However, even in a case such that of a semi-syntactic behaviour in the selection of the different forms of the paradigm of *ise* in Estonian, we do not recognize any striking evidence of a correlation between the use of different $_{CC}$markers with the different uses of intensifiers. Rather, as we have seen, the use of different forms depends on the different forms governed by the bases with which the $_{CC}$marker combines.

If Estonian and Icelandic have different $_{CC}$markers that correspond to different forms of the paradigm of their respective intensifiers (or reflexive pronouns), English and Czech have instead two etymologically distinct $_{CC}$markers which operate within the domain of ACTIONS. Unlike in other European languages, in English the neoclassical combining form *auto-* has a limited productivity in a field, such as the technical-sci-

[132] In all languages of the Finnic branch of Uralic, the object of an action resorts to the partitive case when at least one of the following conditions is not met: 1. the sentence is affirmative, 2. the action is completed, therefore it is resultative, 3. the action expressed by the predicate includes the object in its entirety, or includes a defined part of it. On the other hand, when all three of the above conditions are satisfied, the object is expressed in the genitive case in the singular and in the nominative in the plural (Laanest 1982: 296).

entific one, where this language also plays a prominent role as a source of borrowings for other languages. In fact in English we can find some formations – in addition to the pool of neoclassical formations featuring AUTO- shared by all European languages – such as the IT terms *auto·config(ure)*, *auto·play*, *auto·exec(ute)*. These formations generally indicate functions that are automatically performed by certain technological objects such as cameras, computers, electronic media readers, etc. From a formal point of view, these formations have action nominals or verbal bases as their basis, but their value as automatic functions of objects has lead to these terms being often understood as nouns of (abstract) instruments. In these formations, the function of *auto-* appears limited to the exclusive adverbial function proposed by König (2011), which can be defined, as already said, as subject-oriented. A possible explanation of this functional constraint can be linked to an understanding of the element *auto-* in these formations as a shortening of *automatic(ally)*, an adjective (or adverb) that performs the function of excluding participants alternative to the subject of an action in a way that corresponds to the function of the adverbial intensifier *by X-self*.

Turning our attention to Czech, we find two native markers that can be traced back to two different sources: for *sam-(o-)* the intensifier, for *sebe-* the stressed reflexive. In this case as well there is some competition between the two markers which – in the Czech-German dictionary consulted (czB2) – show a comparable number of different formations (about 50 headwords with *sam-(o-)*, about 40 with *sebe-*). Running through the list of formations with *sebe-*, it is possible to encounter various formations characterized by co-reference between arguments – *sebe·ovládání* 'self-mastery, [Selbstbeherrschung]', *sebe·zničení* 'self-annihilation, [Selbstvernichtung]' – while it is difficult to identify good examples of a subject-oriented interpretation. Among the formations with *sam-(o-)*, on the other hand, we have found both examples of co-reference (*samo·správnost* 'self-administration, [Selbstverwaltung]'), and abundant examples of subject-oriented interpretation with nouns (*samo·regulace* 'self-regulation, [Selbstregulation]') and eventive adjectives (*samo·opalovací (krém)* 'self-tanning cream, [Selbstbräunerkreme]').

The case of Czech provides an example of a language in which the functions proposed in König (2011) have a formal counterpart. However, it should be noted that these examples do not change anything with respect to the possibility of many formations to overturn object-oriented interpretation into subject-oriented. For example, in the case of *sebe·obrana* 'self-defense, [Selbstverteidigung]' – which can have an object-oriented interpretation of the verb *bránit* 'to defend': 'I defend myself [not someone else]'

– it should be verified whether in the case of a subject-oriented interpretation ('I myself defend myself [and no one else defends me]'), it is necessary to use the element *sam-(o-)* instead of *sebe-*. Rather, it is probable that *sebe-*, which, it should be remembered, corresponds to the stressed reflexive pronoun, selects formations where an evident relationship of co-reference can be established between arguments or participants in the action. On the contrary, *sam-(o-)* – consistently with its source, the intensifier *sam* – would be closer to the other pole of contrastive coreference which is represented by the exclusion of alternative agents and external contributions to the action.

We have so far considered examples of languages in which different $_{CC}$markers compete in the area of the conceptual space of POSs characterized by the feature ACTION. Two cases – those of Icelandic and Estonian – show the occurrence of different forms of the same syntactically free element used in compounding, while two other cases – those of English and Czech –show instead etymologically different markers. The type of functional differentiation observed seems to contrast either eventive adjectives to action nominals (s. Icelandic and Estonian) or the pole of the co-referential function (i.e. the use of the adnominal intensifier in a reflexive construction), as opposed to that of the contrast and exclusion of alternative participants (i.e. the use of the adverbial intensifier; s. English and Czech). These tendencies can be summarized in (136) as follows:

(136) **Tendency III.a**

If a language has more than one $_{CC}$marker used with bases lying in the domain of ACTIONS, then the marker which corresponds to the syntactically free element closer to the uses of intensifiers:

a. will occur with deverbal eventive adjectives,

b. will at least have the contrastive function of the exclusion of alternative participants.

This tendency is mirrored by the one in (137):

(137) **Tendency III.b**

If a language has more than one $_{CC}$marker used with bases lying in the domain of ACTIONS, then the marker which corresponds to the syntactically free element closer to the uses of reflexives will at least have the function of defining the co-reference between arguments/participants of the action.

5.2.2 Prototypical verbs (PREDICATION × ACTION)

From the discussion of Figure 18, it has been noted that in some languages $_{CC}$markers are not compatible with bases belonging to the PREDICATION × ACTION area, while the $_{CC}$markers of other languages connect to bases belonging to this part of the conceptual space of POSs. What should be also noted is that, at any rate, all $_{CC}$markers which extend over the PREDICATION × ACTION area also have bases in the MODIFICATION × ACTION area, while $_{CC}$markers which do not extend over the latter (in particular lacking eventive deverbal adjectives) do not reach the PREDICATION × ACTION area. This leads to the formulation of the generalization in (138).

(138) **Generalization II**
In formations with a $_{CC}$marker, the presence of verbs in predicative function implies the presence of deverbal (eventive) adjectives.

In (139), we can propose a tendency that links how close a $_{CC}$marker is to derivation and its use with verbs in predicative function.

(139) **Tendency IV**
The closer a $_{CC}$marker to derivation, the more probable the use of this $_{CC}$marker extends to the use with verbs in predicative function.

This tendency is rather weak for two reasons. On the one hand, we have languages in which a $_{CC}$marker is close to compounding – such as the outcomes of Germanic **selba* especially in Swedish and Danish – but may occur in verbal formations in predicative use. On the other hand, there are several markers which are close to derivation – such as the Hungarian *ön-*, whose source has developed functions that go beyond the intensifiers-reflexive continuum – and exclude these uses. In order to achieve greater generality, we can tentatively modify the formulation in (139) as the generalization in (140) below:

(140) **Generalization III**
If a $_{CC}$marker can be considered a prefix, then it will be used with verbs in predicative function.

In fact, this generalization is also problematic, because if it is possible to propose a relative arrangement of elements of different languages along a continuum between compounding and derivation, it is less obvious to define where compounding is certainly out of discussion and a marker can be surely attributed to derivation. Moreover, placing such a boundary naturally goes against the very concept of continuum.

To clarify the reasons that have brought us to this generalization, we can take into account the examples that support it. The $_{CC}$markers that have so far been considered closest to derivation are AUTO- in Romance languages and *em-* in Breton. In the case of AUTO-, what is crucial is the complete semantic unrelatedness between the $_{CC}$marker and its syntactically free counterpart (AUTO 'car'). The absence of a semantic relationship between them, which also points toward a grammatical function of AUTO-, makes the process closer to derivation. As for Breton, on the other hand, there is a process characterized by features of prototypical prefixation. This element is in fact rigidly in left position in the formations where it occurs, it is etymologically connected to a preposition, it is monosyllabic, and it generates phonological changes in the base that obscure the morphological boundary (i.e. it is a cohering prefix). What is problematic is to provide criteria for excluding elements such as *afto-* in Greek or *ön-* in Hungarian from being as close as AUTO- or *em-* to prototypical prefixation. It should be noted, however, that *afto-* productively occurs with verbs in predicative function.

The morpheme *afto-* is close to compounding because of its formal identity with the personal/demonstrative pronoun *aftó,* and *ön-* presents formal identity with the pronoun of formal address *Ön*. On the other hand, the unrelatedness (in synchrony) of the functions of these syntactically free elements with the continuum of intensifiers and reflexives makes them closer to derivation. By carefully observing what actually distinguishes the cases of Greek and Hungarian from that of Romance languages on the one hand and of Breton on the other, we can indeed find a difference that can perhaps be crucial. The syntactically free elements corresponding to the $_{CC}$markers in Greek and Hungarian do not appear as such in reflexive constructions, and do not constitute intensifiers by themselves. However, the reflexive form of Greek *ton eaftó tou* is recognizably formed by an outcome of Ancient Greek *autós*, while Hungarian *ön-* can appear in a strengthened form of the reflexive construction (*Írország gazdasága (ön) magát pusztítja el* 'the economy of Ireland falls apart/ (destroys itself)'). On the contrary, in Romance languages and Breton even this partial identity is excluded, strengthening the conclusion that the $_{CC}$markers of these languages are derivational units. Finally, as far as Hungarian is concerned, we have

seen in Chapter 3, section 3.2.3, how the class of prefixes in this language is restricted to verbal preverbs, so, differently from Romance languages and Breton $_{CC}$markers, *ön-* would be isolated in a supposed class of Hungarian prefixes.[133]

5.2.3 Semantic classes of the base words

In this section, we will turn our attention to the areas of the conceptual space of POSs of PROPERTIES and OBJECTS. We will show how the functional specialization and the availability of $_{CC}$markers in different languages is affected by bases belonging to these semantic classes. In particular, we will show how in relation to the class of objects, many secondary markers can be identified for the languages of the sample. These secondary markers often coincide with (reflexive) possessives or attributive intensifiers. It will also be observed that under these two conditions – bases belonging to REFERENCE × ACTION; $_{CC}$markers corresponding to (reflexive) possessives or attributive intensifiers – $_{CC}$markers have an attributive/possessive interpretation.

As noted multiple times, not all the languages in the sample have more than one $_{CC}$marker. Languages with a wider inventory of $_{CC}$markers – for example, Slavic languages and especially Czech – contrast with languages characterized by a single $_{CC}$marker, such as Maltese and Romance languages. The peculiarity of the languages just mentioned lies in the fact that, in addition to having a single $_{CC}$marker, these markers share their etymological source with the $_{CC}$marker of Modern Greek, which in turn goes back to the $_{CC}$marker attested in Ancient Greek. If we compare the compatibility of these three $_{CC}$markers with different bases, we obtain the semantic map in Figure 20.

It can be noted that Modern Greek *afto-* is the $_{CC}$marker with the widest compatibility in the conceptual space of POSs, being compatible with bases belonging to the class of ACTIONS (*afto·dhiapséudhomai* 'to contradict oneself' PREDICATION × ACTION), to the class of PROPERTIES (*afto·dhínamos* 'self-sufficient' MODIFICATION × PROPERTY) and of OBJECTS (*aftó·phōtos* 'that shines with its own light', lit. 'self-light' MODIFICATION × OBJECTS). The marker AUTO- of Romance languages, on the other hand, is compatible with ACTIONS and to a lesser extent to PROPERTIES (Spanish *auto·confianza* 'self-confidence', French *auto·suffisance* 'self-sufficiency', both REF-

[133] See also Kiefer (2016: 3309): «Derivation is performed exclusively by suffixation. The language does not distinguish between particle and prefixed verbs, all complex verbs are particle verbs. The prefix *leg-* marks the superlative, but comparison does not belong to word-formation.»

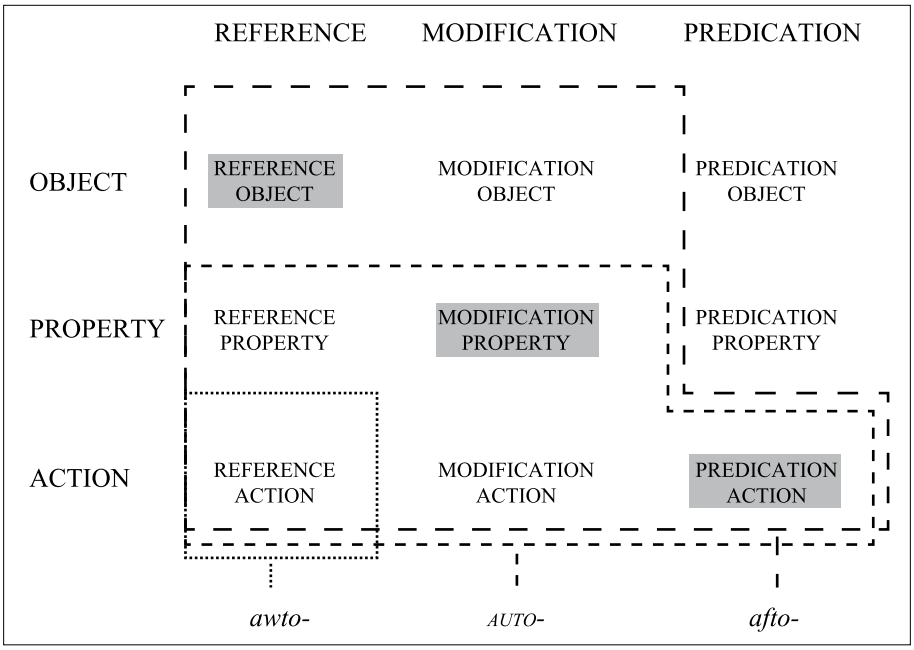


Figure 20. Semantic Map (3): $_{CC}$Markers developed from Ancient Greek *auto-* and their compatibility with POSs

ERENCE × PROPERTY / STATE). The case of Maltese has already been discussed above.

For other European languages, this type of phenomenon, i.e. the different compatibility of markers of different languages sharing the same etymological origin, is not as clearly visible as in the case of the various outcomes of Ancient Greek *auto-*.[134] The markers cognate to Germanic **selba* or the $_{CC}$markers of the Slavic languages of the SAM-(O-) type appear to have a compatibility with different areas of the conceptual space of POSs similar to that of AUTO- in Romance languages, with the exception of the differences in compatibility in the area of PREDICATION × ACTION. Conversely, an extreme case is that of the group of Baltic languages in the Indo-European language family. In both Latvian and Lithuanian only one $_{CC}$marker is available, but the syntactically free elements which the two languages draw on are different. This is actually striking if we consider that the two languages use cognate markers (see Chapter 4, section 4.2.5) for similar functions in syntax. We can summarize these observations in the generalization in (141).

[134] A possible explanation of the reasons for this difference is put forward in 6.2.4 below.

(141) **Generalization IV**

Etymologically related $_{CC}$markers in different languages do not necessarily display the same compatibility with bases in the conceptual space of POSs.

Returning to Figure 20, the compatibility with the semantic classes of words that characterizes Modern Greek *afto-*, as represented by the map, is analogous to the compatibility observed for Albanian *vet(ë)-* and for Hungarian *ön-* (with the exception of the area of PREDICATION × ACTION in the latter). Also, those are the only $_{CC}$markers available in these languages: this is especially true for Albanian, which does not appear to have paradigmatic alternatives for expressing contrastive coreference, while in Modern Greek and Hungarian the $_{CC}$markers *afto-* and *ön-* are flanked by *idhio-* and *saját-*, respectively.[135]

It has already been mentioned though that other languages in our sample have more than one $_{CC}$marker. We have dealt with some cases of competition between different markers in 3.2.1 above, in particular for the area of ACTIONS. We will now broaden our scrutiny to include other semantic classes and other markers.

Germanic languages, Slavic languages and Balto-Finnic languages of the Uralic family (Finnish and Estonian) show the existence of a "main" series of $_{CC}$markers that extend over the area of ACTIONS (the various outcomes of **selba*; SAM-(O-), Czech *sebe-* and Russian *sebja-*; Finnish *itse-* and the various forms of Estonian *ise-*), and of another "secondary" series that extends over the area of OBJECTS (German *eigen* and cognate elements; the various markers corresponding to the reflexive possessive and attributive intensifiers in Slavic languages; Finnish and Estonian *oma-*).

Figure 21 shows the extension of these series of markers to the conceptual space of POSs. The area of compatibility of most $_{CC}$markers is defined by the gray area,[136] while the second series of markers is defined by different types of broken lines defining areas of different size.

The distinction between main and secondary $_{CC}$markers depends on an evaluation of their productivity: the series of formations with main $_{CC}$markers – whose

[135] The latter marker corresponds to the attributive/possessive intensifier – but to our knowledge, there is competition only in one formation: *saját·kezű(leg)* vs. *ön·kezű(leg)* 'by one's own hand' – while the former corresponds to the current adnominal intensifier (PRO/NP ART *ídhi-*, e.g. *eghó o ídhios* 'myself' TDIR) and differs instead from the attributive intensifier (*dhikó* 'own' TDIR).

[136] The area included in the dotted line shows the compatibility of some $_{CC}$markers with bases in PREDICATION × ACTION.

productivity is largely ensured by bases belonging to the class of ACTIONS – are more numerous than the series formed with secondary $_{CC}$markers. The latter combine with bases belonging to the class of OBJECTS, but also with bases belonging to the class of PROPERTIES/STATES. The widest compatibility for secondary $_{CC}$markers is attested in the case of Germanic and Finnish and Estonian languages: in these languages, some formations with secondary $_{CC}$markers display bases belonging to the area REFERENCE × ACTION.

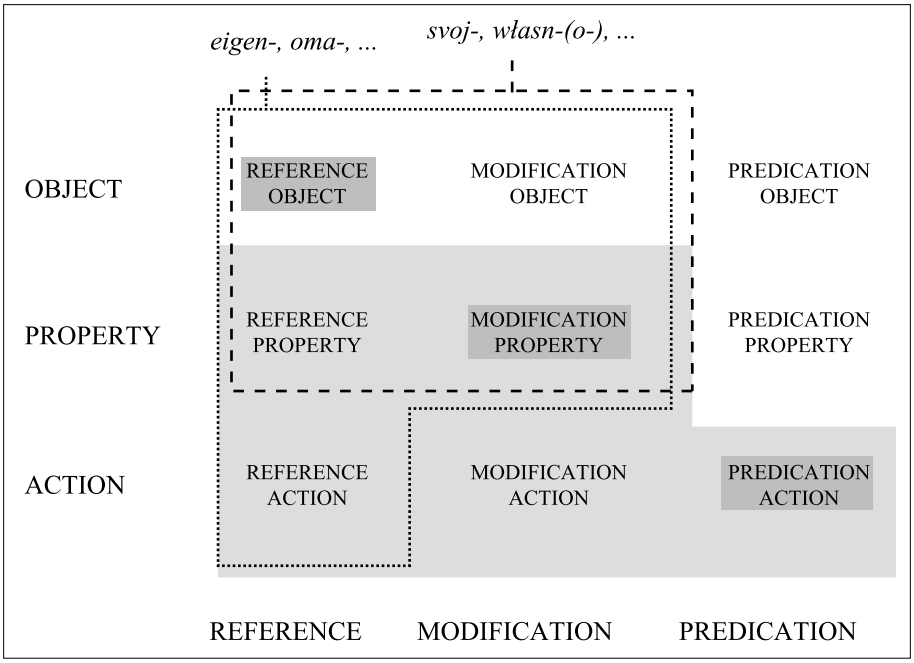


FIGURE 21. Semantic Map (4): Functional specialization of $_{CP}$markers

All formations with a secondary $_{CC}$marker in Slavic languages and most of those in Germanic and Finnic languages have an attributive-possessive interpretation if the respective bases belong to the classes of OBJECTS and PROPERTIES/STATES. Formations such as Danish *egen·produktion* 'own production', or Estonian *oma·toodang* 'own production', which display a REFERENCE × ACTION base, can still be interpreted as having an attributive-possessive interpretation, probably linked to the shift of the action nominal from a process to a result interpretation (*production* 'process of producing' → 'result of producing/product(s)'). German *Eigen·transplantation* 'autotransplant' or Finnish *oma·tekoinen* 'self-made' appear instead closer to the

interpretation conveyed by the primary $_{CC}$markers *selbst-* and *itse-*. This does not, however, exclude that German and Finnish secondary markers can have an attributive-possessive interpretation also with REFERENCE × ACTION bases – compare *Eigen·bedarf* 'own use' or *oma·tuotanto* 'own production, [Eigenproduktion]'.

Consequently, the functional specialization that characterizes many secondary $_{CC}$markers is that of having an attributive/possessive interpretation by which we contrast the actual owner of an object and exclude alternative owners. The competition between main and secondary $_{CC}$markers has the effect of constraining the compatibility of main $_{CC}$markers to bases belonging to the class of ACTIONS, and excluding their use with OBJECTS and thus from the attributive/possessive interpretation. We can define this interpretation as **contrastive possession** which expresses that the OBJECT constituting the referent of the base of the formation is in relation with a possessor, with whom a personal and peculiar possession is established.

On the basis of the different compatibility of the two series of brands with the POS, the tendency in (142) can be proposed.

(142) **Tendency V**
If a language has more than one $_{CC}$marker, a functional specialization will be observed, shifting from the class of ACTIONS to the class of OBJECTS.

This tendency may result in a more predictive generalization, such as the one proposed in (143).

(143) **Generalizations V**
If a language has more than one $_{CC}$marker and one of them extends to the class of OBJECTS, then:
a. this marker will not be the main $_{CC}$marker,
b. the main $_{CC}$marker will not be compatible with bases belonging to the class of OBJECTS, and in any case it will not have an attributive/possessive interpretation.

The case of the relationship between *ön-* and *saját-* seen above constitutes a counterexample to the generalization which, however, does not undermine the general tendency towards the functional specialization of different markers either on the class of OBJECTS or on that of ACTIONS.

In the case of markers such as German *eigen-* or Finnish and Estonian *oma-*, one can still speak of $_{CC}$markers extended over a wider territory than that of the main $_{CC}$markers, and with an interpretation of contrastive possession. However, in the case of markers such as those of Slavic languages, in which the relationship between the function of contrastive possession and certain markers is biunivocal, we can propose the label of $_{CP}$markers, that is, markers of contrastive possession.

5.3 Conclusion: Constructional schemas and functional specialization of contrastive coreference markers

In Chapter 4, we have identified four different types of $_{CC}$markers at the borderline between derivation and compounding. In the survey of the compatibility of $_{CC}$markers with different kinds of bases, we have observed that some $_{CC}$markers are compatible with bases that lie outside the domain of ACTIONS, which constitutes the semantic class with which $_{CC}$markers are most productive. This was not unexpected since we were aware of the fact that in German and other languages we have a competition between different markers. However, what we have shown in this chapter is that some $_{CC}$markers have a wider compatibility with other semantic classes of words. That is, in some languages some markers are polysemous, and can be regarded as $_{CC}$markers and $_{PC}$markers at the same time.

This case can be considered as a further type of compound $_{CC}$marker, represented in Figure 22 by using the representation of lexical hierarchies used in Construction Morphology.

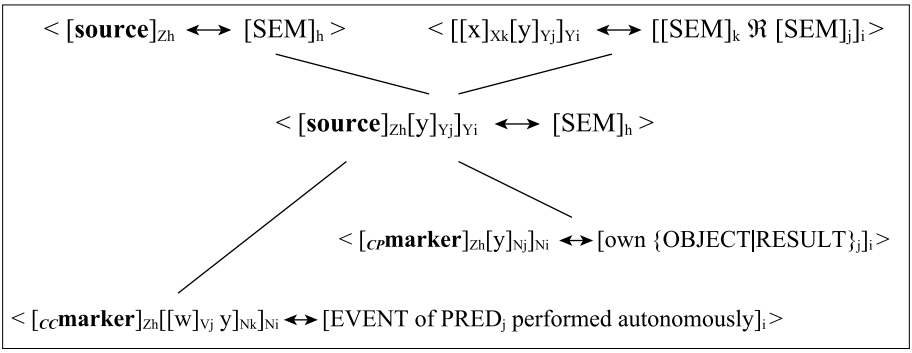


Figure 22. Compound $_{CP}$markers I

Modern Greek *afto-* and Hungarian *ön-* are examples of this schema. In Chapter 4, we attributed them to the type "Compound $_{CC}$marker III", but it is not to be excluded that other types of compound markers can convey the function of contrastive possession as well. At any rate, the case in which a marker has the function of both a $_{CC}$marker and a $_{CP}$marker can be considered as a first type of compound $_{CP}$marker. As for Lithuanian *sav-(a-)*, which differs from the $_{CC}$marker *sav-(i-)* because of the different linking vowel, it is rather a case of competition of a $_{CC}$marker with a $_{CP}$marker.

In this more usual case, $_{CP}$markers can be represented as in Figure 23. In the data, we have seen that many $_{CP}$markers have their source in attributive intensifiers, which can be categorized as adjectives.

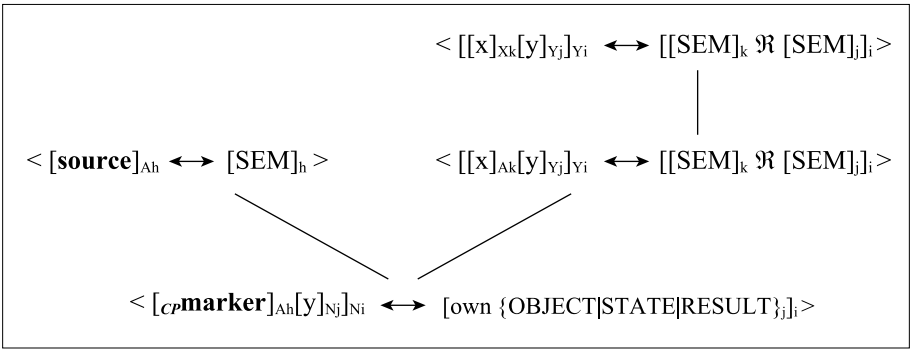


Figure 23. Compound $_{CP}$markers II

In this case, it is not necessary to refer to the syntactic construction of the attributive intensifier through a paradigmatic relation, because that function is performed by an adjective meaning 'own', and so its contribution to the compound is rather compositional.

The synchronic survey of $_{CC}$markers carried out in this and the previous chapter has exposed some limits as far as the compositionality, or more in general, the motivation of $_{CC}$formations is concerned. If the cases of derivational $_{CC}$markers do not posit a major problem, since derivational schemas are by definition conventionalized form-meaning pairings, the case of compound $_{CC}$markers closer to derivation, i.e. less motivated and weakly compositional, constitutes a challenge. In Chapter 6, we will attempt to account for such gaps in motivation networks by resorting to diachronic or contact explanations.

6

DIACHRONY AND CONTACT

In the previous chapters, we have outlined a synchronic picture of $_{CC}$markers in the languages of Europe. From the analysis of the data and the trends observed, we have concluded that most of the current European languages express contrastive coreference in word-formation through the markers obtained by drawing mainly on the pool of syntactically free elements that belong to the intensifiers-reflexive continuum. With regard to the compatibility of these markers with specific word classes, we have observed that for the most part, the $_{CC}$markers having intensifiers as their source are compatible with bases belonging to the class of ACTIONS, and to a lesser extent to that of PROPERTIES. A second area of compatibility of $_{CC}$markers concerns OBJECTS. In some European languages, formations compatible with the class of OBJECTS have a specific marker that is usually derived from either the (reflexive) possessive or the attributive intensifier. In these cases, this specific marker also corresponds to an interpretation different from that of contrastive coreference, which we have defined as contrastive possession (*CP*): in these cases, the markers specialized in this interpretation can be defined as $_{CP}$markers.

The trends observed in the previous chapters and summarized above characterize many, but not all European languages of today. Some deviations from these trends concern the source, but most of all the type of procedure that characterizes a $_{CC}$marker. Other deviations from the general trends, however, concern the compatibility between markers and word classes.

In the first part of this chapter, we will consider some deviations from the major trends observed in Chapters 4 and 5 and will propose explanations for these deviations, which have been derived from considerations concerning diachronic developments or language contact among European languages (6.1).

One of the most remarkable deviations from the trends observed for European languages concerns AUTO- in Romance languages. The set of etymologically related markers in these languages belongs to the domain of derivation – therefore, they are contrasted with the tendency of European languages by which $_{CC}$markers constitute units of compounding. Their development in Romance languages, however, is intertwined with that of other European $_{CC}$markers that belong to compounding. The overview of this development will give us the opportunity, in the second part of the chapter (6.2), to observe how linguistic contact phenomena have repeatedly characterized the diffusion, productivity and restrictions of $_{CC}$markers in the languages of Europe. In the final section (6.3), we will again

resort to Construction Morphology to model the developments observed by using its formalism.

The overview presented in this chapter cannot and does not aim to be exhaustive. Rather, it is intended as a repertoire of suggestions for further research in the field of the diachrony and contact of $_{CC}$markers in European languages. In addition, the aim of this chapter is to show how contact phenomena concerning the lexicon can be limited to a simple loan or calque of individual foreign formations, but can also induce the development of productive and *de facto* nativized procedures in a language.

6.1 Exceptions to the tendencies

In this paragraph, we will take back into consideration the cases of those $_{CC}$markers which constitute exceptions to the tendencies observed in Chapters 4 and 5, and we will suggest diachronic explanations of these deviations. The exceptions observed can be grouped into two types: a first type concerns markers whose corresponding syntactically free element does not belong to the intensifiers-reflexive continuum (6.1.1). A second type concerns markers which lack a corresponding syntactically free element (6.1.2).

6.1.1 Sources outside the intensifiers-reflexive continuum

In Chapter 4, we have observed that most of the $_{CC}$markers in present-day European languages have as their source syntactically free elements that perform one or more of the functions constituting the intensifiers-reflexive continuum. It has already been noted above, however, that some of the $_{CC}$markers identified in Chapter 4 correspond to elements that to varying degrees do not belong to the intensifiers-reflexive continuum. The $_{CC}$markers in question are Hungarian *ön-*, Turkish *öz-* and Modern Greek *afto-*.

As for the Hungarian $_{CC}$marker, a clarification must be made. In Hungarian, as already seen in Chapter 4, section 4.2.7.2 above, *ön* can occur in the reflexive construction to strengthen *maga*, and is rigidly placed before it. Moreover, it cannot function as the intensifier on its own, or as the reflexive without *maga*, which is instead the main intensifier in Hungarian today. On the other hand, as

a syntactically free element, *ön* is a personal pronoun with social deictic functions. In this function, *ön* again competes with *maga*, which is also a pronoun of formal address, although *ön* is the more formal or polite of the two (Kenesei *et al.* 1998: 266-267). What is relevant in this chapter is to note the discrepancy between the function of *ön* in its use as a syntactically free element and as a $_{CC}$marker. Given also the scarce formal independence and the absence of a semantically specific contribution of *ön* within the reflexive construction, if it may seem arbitrary not to relate this function to that of the $_{CC}$marker, it seems to us equally arbitrary to do the opposite and make the two functions depend on one another. In terms of diachronic development, the occurrence of *ön* in these three constructions – in two of which it occurs as a bound element combined with *maga* – can be explained by the fact that *ön* originally had the function of intensifier, reflexive and emphatic possessive, combining the current functions of *maga* and *saját*. This is the situation found in the attestations of *ön* from the fourteenth and fifteenth centuries (Benkő: 27). Starting from these functions, in Hungarian we are witnessing on the one hand the development of a new marker of intensification and reflexivity (*maga*), while the old marker *ön* continues to occur bound in one competing construction (*önmaga*). On the other hand, *ön* evolves from the intensifier to a pronoun, up to the current function of a pronoun of formal address.

As far as the $_{CC}$marker *ön-* is concerned, its attestation dates back to the 18th century in formations comparable to German compounds with *selbst-* – for example, *ön·alló* (≈ *selb(st)·ständig*) is attested starting from 1830. This is the time in which Hungarian culture experienced a nationalistic ferment and the lexicon – previously enriched mainly through loans from German, due to the ever growing cultural influence from Austria – started to grow thank to the coinage of calques, especially as regards the philosophical-scientific lexicon (Benkő 1972: 184-185). It is not possible to establish whether the use of this marker derives from a group of formations already available in Hungarian as a model at the time when the calques from German were coined. Also, it should be ascertained since when *ön* was no longer available for functions related to the intensifiers-reflexive continuum. A more accurate analysis of intensifiers in the time of attestation of the first formations with $_{CC}$markers could shed light on the reasons at the core of the current situation, which synchronically contrasts with the tendencies observable for most European languages.

A situation quite similar to that of Hungarian occurs in Modern Greek. In this language, the syntactically free element corresponding to the $_{CC}$marker no longer displays functions linked to the intensifiers-reflexive continuum. Like Hungarian *ön*, Modern Greek *aftó* has a place within the pronominal system as the emphatic pronoun *aftó*, which occurs as a $_{CC}$marker in form of a bound morpheme, and also in the reflexive construction as part of the pronoun *eaftó* (see below in 6.2.1 for the origin of this form). Similarly to Hungarian, within the reflexive construction it is not possible to distinguish the specific contribution of *aftó,* which, moreover, from a formal point of view, can never be separated from *eaftó*, an element that constitutes a syntactically indivisible unity within a rigid syntactic construction (see section 4.2.6.1 and example (102) above).

For Modern Greek, the use of *aftó* as a source of the $_{CC}$marker can be explained by the substantial continuity between the lexicon of Ancient Greek and that of Modern Greek, especially as far as the written language and its impact on the spoken variety and on the philosophical and scientific lexicon are concerned.[137] A more in-depth survey of the origins, development and conditions of maintenance of formations with *afto-* along the history of Greek will be pursued in the second part of the chapter (4.2).

The situation of Greek and Hungarian can therefore be summarized in the diagram in Figure 24. The diagram a. represents the diachronic evolution of Greek markers, while diagram b. represents that of Hungarian ones. The elements on the right of each diagram – which represent the current functions, performed by syntactically free elements or bound ones – are diachronically linked to the element on the left, which represents the historical antecedent belonging to the intensifiers-reflexive continuum. The grey arrows from left to right indicate a relationship of diachronic development. Beyond this diachronic relationship, we believe that there are no synchronic relationships between the various forms from the semantic point of view.[138]

[137] «[A]nother major difference between A[ncient]G[reek] and M[odern]G[reek] word-formation is caused by the influence of the diglossic history of the language. All historical phases of Greek, from the Koine onwards, display a distinction between learned/high and vernacular/low which widens with the passage of time, involving more and more phonetic and morphological differentiation. As a result, the vocabulary contains lexical items of different chronological periods and different morphological properties.» (Manolessou and Ralli 2016: 2043-2044)

[138] We will formalize this representation in section 6.3 below by using the formalism of Construction Morphology.

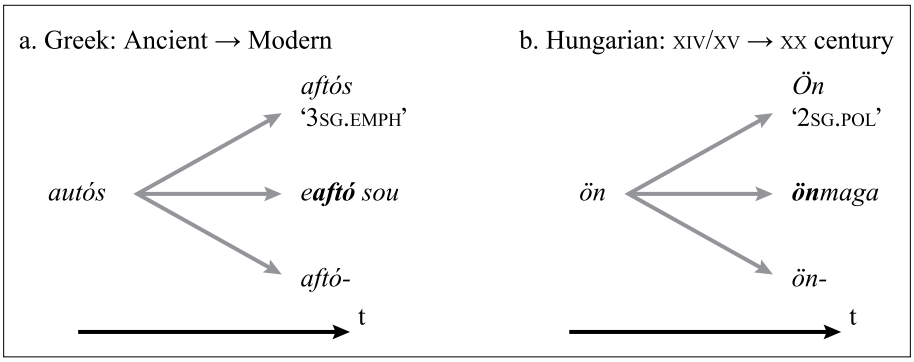


Figure 24. Intensifiers, reflexives and $_{CC}$markers across time (1): Greek and Hungarian

Another language in our sample, i.e. English, displays a situation similar to that of Greek and Hungarian, though it also shows some noteworthy differences. In English, the $_{CC}$marker *self-* corresponds, strictly speaking, to an element that belongs to the intensifiers-reflexive continuum. It is, however, disputable whether this element is syntactically free, given that it is always linked to forms of the paradigm of personal pronouns. In Chapter 4, sections 4.2.2 and 4.3, the presence of *self* as a noun with the meaning of 'a person's essential being that distinguishes them from others' (OED) – therefore belonging to the pool of lexical sources of intensifiers – was considered, from a synchronic point of view, as a sufficient condition to consider the noun *self* and the intensifier-reflexive PRO-*self* together as possible sources of the $_{CC}$marker, similarly to other Germanic languages. Given the diachronic perspective of this chapter, it should be noted that the noun *self* is not the source of the intensifier, but rather the opposite, as the current compound pronoun PRO-*self* is a univerbation given by the Old English intensifier *self* (alternative forms: *silf*, *sylf*, *seolf*) preceded by the dative or accusative forms of the personal pronoun, some of which undergo a phonological process of reduction of the vowels in unstressed syllables (for example, *myself* < *me self*).[139] The intensifier *self* in Old English, when combined with object personal pronouns, effectively disambiguated the interpretation of 3SG pronouns, which could otherwise be interpreted both as co-referent with a preceding subject or not.

[139] Compare König and Siemund (2000: 46) and the literature discussed for more details and alternative hypotheses.

(144) Old English (Ælfric of Eynsham, *Letter to Sigeweard* 1129; cited in König and Siemund 2000: 45)
se Hælende sealde hine sylfne for us
'The Saviour gave himself for us'

These compound forms of the reflexive pronoun have been used from the Early Middle English period as intensifiers in place of the simple *self* (König and Siemund 2000: 46). From the seventeenth century, the noun *(the) self* has developed as a substantivization of the intensifier and with the sense of 'what is typical and intrinsic of a person'. This series of developments is summarized in Figure 25.

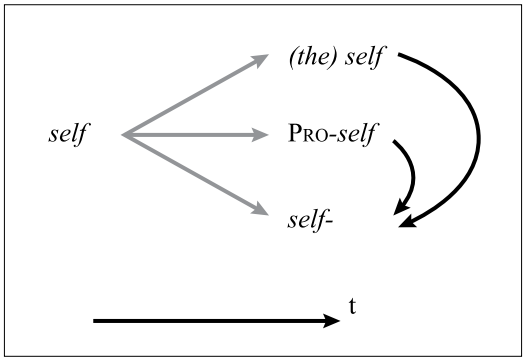


FIGURE 25. Intensifiers, reflexives and $_{CC}$markers across time (2): English

The black arrows indicate that, unlike in the case of Hungarian and Modern Greek, it can be assumed that the noun *(the) self* can be in synchrony interpreted by speakers as the source of *self-* as a unit in compounding.[140] Furthermore, a formal similarity between the possessive pronoun and the first element of the compound reflexive pronoun is still apparent today, and therefore allows the latter to be synchronously analyzed as a sequence of a possessive and the noun *(the) self*.

The situation is rather different in Turkish. In this language, the $_{CC}$marker *öz-* has a syntactically free counterpart that can occur as an adjective or a noun. The meaning of the noun ('essence, self') points to its inclusion within the pool of lexical sources of intensifiers. However, *öz* does not display any grammatical function in Turkish on the intensifier-reflexive continuum.

[140] See the schema of compound $_{CC}$markers of type I in Chapter 4, section 4.4.

About this anomaly, some considerations must be made. The lexicon of today's Turkish is the result of a number of policies and reforms which have affected the new Turkish state that arose in the Anatolian peninsula in the last century, starting from the decline of the Ottoman Empire at the end of the First World War. First of all, a reform of Ottoman Turkish took place following a puristic orientation aimed at the removal of (almost) all words of Arabic and Persian origin, and with the intent of a cultural democratization of the lexicon: the Arab-Persian layer was considered a factor that created a gap between the language of culture and the language of people. In Lewis' (1999) words:

> «The mixture of Turkish, Arabic, and Persian, which Turks call Osmanlıca and we call Ottoman, was an administrative and literary language, and ordinary people must have been at a loss when they came into contact with officials.» (Lewis 1999: 8)

In addition to the claim of democratizing the language, a nationalistic intent was also driving the language reform. This intent provided the method for carrying out the reform: the lexicon was artificially enriched with words formed with roots and morphological processes considered as peculiarly Turkish.[141] The sense of Turkish purity is also manifested by the name *Öztürkçe* ('pure Turkish', which includes the adjectival use of *öz* 'pure') that designates Reformed Turkish, as opposed to *Osmanlı Türkçe* 'Ottoman Turkish'. After the Second World War, a wave of borrowings, mostly from English, enriched the Turkish lexicon – a process shared by many other European languages.

The coinage of dozens of words is the effect of the Turkish linguistic reform. From *öz* have been derived for example: *öze*/*özgü* 'peculiar (of)', *özgür* 'free', *özek* 'center', *özel* 'private', *özerk* 'autonomous', *özet* 'summary', *özgül* 'specific', *özgün* coined to mean 'original', but used as 'authentic'. Some derivatives, on the other hand, pre-date the reform: *özen* 'care, attention', *özge*

[141] Some newly introduced processes are not actually attested in the Turkish tradition, but are ideologically considered Turkish on the basis of a theory (*Güneş-Dil Teorisi* 'theory of the language of the sun') – encouraged for a certain period by Mustafa Kemal Atatürk in person – who argued, among other things, that Turkish was the first language to take shape (an idea inspired by the theories exposed in *La Psychologie de quelques éléments des langues turques*, by Viennese Hermann F. Kvergić), and that all languages, including those recognized as non-Turkish, are ultimately derived from Turkish (Lewis 1999: 57-58).

'other', *özenti* 'emulation' (Lewis 1999: 144). The terms introduced by the reform may even involve elements of word-formation of foreign origin – *özel* is formed with the suffix *-(s)el/-(s)al*, which actually issued from French: compare *cultur·el* 'cultural' – or derive from words and processes of Turkish origin, which, however, display a non-compositional and arbitrary meaning – e.g. *özgürlük*, formed from *öz* 'self' and *gür* 'abundant', means 'freedom', although from its constituent elements it should mean 'abundant of self, of essence'. This is not a case of semantic drift, but of arbitrary coinage, which also involves the aforementioned *özgür*, almost certainly a backformation from the noun (Lewis 1999: 120).

The formations with *öz-* as a $_{CC}$marker certainly do not display a non-compositional meaning; however, they are most probably recent consequences of the Turkish pressure to introduce purely Turkish neologisms in terminology. For example, the dictionary of A. Vahid Moran in its 1945 edition (tkB1) reports – in addition to some of the formations with *öz* listed above – only *öz·indukleme* 'self-induction' and *öz·saygı* 'self-respect' as formations with $_{CC}$marker. The 1985 edition of Moran's dictionary (tkB2) instead reports a substantial number of other compounds, including *öz·güven* 'self-confidence', *öz·eğindini* 'auto-suggestion', *öz·kavrama* 'self-concept', *öz·kösnülüğü* 'autoerotism', *öz·veren* 'self-sacrificing', *öz·yanma* 'self-combustion'.

Concerning the use of *öz* as a $_{CC}$marker, the choice of this element over the reflexive intensifier *kendi* could be linked to the fact that in other Turkic languages, words cognate to *öz* are still used as intensifiers today, and can also be found in the languages such as Uzbek: compare *o'zing bilasan* 'you yourself know' (uzB1: 578). In addition, formations with these correspondents of Turkish *öz* are indeed used as $_{CC}$markers, for example, Azerbaijani *özünü·tənqid* 'self-criticism' or Uzbek *o'z·belgilash* 'self-determination'.

At the end of this section, we can observe how the cases of major discrepancies between the trends observed at the end of Chapter 4 for the $_{CC}$markers in the languages of Europe – i.e. the cases of Hungarian, Modern Greek and Turkish – can be reduced to the general tendencies if considered from a diachronic perspective as the permanence in synchrony of markers, which in other phases of the language (or concerning Turkish, in related contemporary varieties) have had syntactically free counterparts regularly belonging to the intensifiers-reflexive continuum.

6.1.2 Absence of syntactically free correspondents

Unlike the markers seen in 6.1.1, in this paragraph we will consider the $_{CC}$markers that lack a corresponding syntactically free element altogether. To this category belong the various outcomes of Ancient Greek *auto-* in Romance languages and Maltese and Breton *em-*.

We have seen in Chapter 4, section 4.3, how the case of AUTO- and Breton *em-* are the clearest cases of derivational processes among $_{CC}$markers, both from a semantic and from a formal point of view. However, the case of AUTO- appears to be a result of contact phenomena: the loan of a series of formations and eventually of a word-formation process from one language (simplifying: Ancient Greek) to another (each of the languages where AUTO- is synchronically productive). As such, AUTO- will therefore be considered more carefully in the second part of this chapter in 6.2.4.[142]

On the contrary, Breton *em-* is a case of the development of a prefix with reflexive function, or, according to our proposal, of contrastive coreference, starting from native material. We can find insights on the origin of this element in historical grammars and dictionaries (Fleuriot 1964; Lewis and Pedersen 1961; Loth 1982). The sources mentioned suggest for *em-* a quite usual pattern for prefixes cross-linguistically, i.e. its evolution from a preposition: the Indo-European locative preposition *ambhi*/*m̥bhi* (POKORNY: 34). Cognate prefixes are also found in preverbs of other Celtic languages, as the Irish Gaelic *imb-* – corresponding to the Irish Gaelic preposition *im*/*imm* 'around' – the Welsh preverbs *am-*/*ym-*, the Cornish preverbs *am-*/*em-*/*ym-,* and again in Breton in the prefix *am-*. In these latter languages, the forms *ym-* and *em-* are used to form reciprocal and reflexive verbs –Middle Welsh *yd* ***ym-****gerynt* 'that they loved each other', Cornish ***em-****lathe y honan* 'to kill himself',[143] Middle Breton ***em-****rentaff* 'se rendre' – while in Irish, it is the form *imm-* which developed the use of preverb of reflexive or reciprocal value – ***imm-****un-cualammar* 'we had heard from one another' (Lewis and Pedersen 1961: 263-264).

[142] From a synchronic point of view, see above Chapter 2, sections 2.2.4-2.2.5, for a discussion of the status of *auto-* in Italian. See Chapter 4, sections 4.2.1 and 4.2.9, for a discussion of the outcomes of the Ancient Greek *auto-* in Romance languages and Maltese respectively.

[143] In the example from Cornish, *honan* is the intensifier cognate to Welsh *hunan*, see Chapter 4, section 4.2.4, above.

In Old Breton, the presence of the preverb was sufficient to convey the value of reflexive and reciprocal, while in Middle Breton, in order to indicate reflexivity, it became necessary to precede the preverb with a pronoun agreeing in person with the subject, but having a non-nominative form. In modern Breton, a further development leads the third person singular masculine pronoun to generalize for all persons, and the original preverb becomes the second element of a complex reflexive marker: *en em* (also in a reduced form *'n'em*/*'n'en*) (Fleuriot 1964: 331-332).

It is not clear how for Breton *em-* and the other similar Celtic preverbs a reflexive and reciprocal meaning develops from the locative meaning 'around', but from a diachronic point of view, the current use as a $_{CC}$marker – which, it should be reminded, is of limited productivity – appears to derive directly from the bound use of the preverb rather than from the subsequent use as a marker of reflexivity (i.e. *en em*). Concerning the development of the reflexive and reciprocal meaning from the locative one, some examples, including that of the Welsh noun *ym·ladd* 'battle', from *lladdaf* 'kill, put to death', would suggest the possibility that it is rather the reciprocal that emerges first, through an interpretation of the preverb *ym-* from 'around' to 'each other'. Further research in this direction might shed light on this development, which, if confirmed, would constitute a counterexample to the usual cline in the development of reciprocal markers from reflexive ones, substantiated by examples in many languages (cf. König and Siemund 1999: 58, Figure 3, 59).

6.2 Spread of contrastive coreference markers: Loanwords, calques and triggers of productivity

This paragraph will outline the spread of formations and $_{CC}$markers from Ancient Greek to the contemporary European languages. The survey will in chronological order touch on Ancient Greek, Latin, ancient attestations of some European languages, the relationship between Romance languages and neoclassical lexicon, and finally, on the possible instances of contact among languages of Eastern and Northern Europe.

Through the path outlined, we aim at providing an overview of the possible mutual influences among European languages with respect to the formations examined in this book. Another aim is to highlight the major and clearest cases of con-

tact or cross-linguistic influence, while also pointing at other plausible instances of contact which should be more carefully investigated in future research. In some cases, we will bring forward the hypothesis that the development of formations with $_{CC}$markers might have emerged independently from contact.

At any rate, in the outline we will show how Ancient Greek and its $_{CC}$marker *auto-* undoubtedly played a central importance against which the autonomy and relevance of other contact phenomena considered will inevitably be compared.

6.2.1 Ancient Greek

In 4.2.1, it has been noted how the $_{CC}$marker AUTO-, originating in Ancient Greek, is synchronically represented and (to varying degrees) productive in various European languages. It is therefore worthwhile to take stock of Ancient Greek and of its $_{CC}$markers.

In Ancient Greek, *auto-* as a unit compounding had its most likely source in the intensifier *autós*, *-tḗ*, *-tó*. Puddu (2005: 206-222) compares the main elements with the function of intensifier in Greek (*autós*) and Latin (*ipse*), and analyzes them in the light of the literature devoted to this functional category.

In the context of traditional grammars, three uses of the pronoun *autós* have been identified:

i. a predicative use with the meaning of 'same', 'in person' (compare Latin *ipse*);
ii. an attributive use with the meaning of 'same' (Latin *idem*);
iii. an absolute use as a third person pronoun (Latin *is*). (Puddu 2005: 207-208)

Puddu notes how the emphatic value of *autós* persisted throughout the classical era in the nominative and as an adjunct, while in the other cases its value is exclusively that of anaphoric pronoun, which anticipates its function as a personal pronoun in Modern Greek (Puddu 2005: 208). The analysis of *autós* on the basis of the dimensions of variation of the category of intensifiers proposed by König (2001), shows that *autós* is distinguished from the reflexive (*heautón*, formed by the Homeric reflexive pronoun *hé* and *autós*), agrees in gender, number and case, but not in person, with the constituent to which it refers, and is characterized by the adverbial (inclusive and exclusive) and the adnominal

functions of intensifiers. It should be added to these features that *autós* can also act as an emphatic possessive (an attributive intensifier in König's 2001 terms) in connection with the possessive pronoun.[144] The possibility of *autós* to act as an attributive intensifier may explain its compatibility in word-formation with the semantic class of OBJECTS in the conceptual map of POSs shown in Chapter 5, section 5.1.[145]

Moving on to the use of *autós* in word-formation, let us start with an analysis of the development of the value of compounds with *auto-* from the archaic to the classical age. Sforza (2007) analyzes the importance and the evolution of the meaning of some formations with *auto-* attested in Homer and in the Greeks tragedians Aeschylus and Sophocles. If in Homer these formations have a particular connection with the sphere of life, a greater connection with death and with reflexive actions that result in bloodshed is attested in the tragedians. Beyond literary interpretations, Sforza's observations show the presence of formations with *auto-* already in the archaic phase of Greek literature, and shows how the subsequent semantic developments of these formations, although involving profound changes of meaning, remain within the scope "of subjectivity, even in the cases in which the subject is not actually affected" (Sforza 2007: 112; our translation). This lack of affectedness, which actually corresponds to the absence of direct reflexivity, can be seen especially in compounds with *auto-* used in tragedians. These compounds do not necessarily express an action perpetrated against oneself (e.g. the killing of oneself, the suicide), but can instead indicate the fact that the action is carried out on the *génos*, that is, on a family sphere closely linked to the subject, and therefore only indirectly affects the subject himself. In later times, then, terms such as *autó·kheir* or *authéntes* (corresponding to the lat. *parricidas*) develop in the language of orators a more generic value of 'material performer' or 'he who acts with his hand', as the former reflexive value disappears (Sforza 2007: ibidem).

While considering examples of various kinds, Sforza limits his conclusions on Ancient Greek to the observation of the functions most closely related to reflex-

[144] The described state of affairs evolves in the attestations of Attic prose: "[...] while in the first and second person emphatic possession is still marked by adding *autoũ*/*autón* to the possessive, in the third person the genitive of the reflexive pronoun [*heautoũ*] is used." (Puddu 2005: 200, our translation)

[145] On the diachronic evolution of this compatibility with objects in the Romance outcomes of Greek *auto-* see 6.2.4 below.

ivity and the involvement of the ego. By broadening our perspective to a wider range of examples and considering the compatibility of *auto-* with the various areas of the conceptual map of POSs, it is possible to identify and exemplify the different functions already spotted in the previous chapter for contemporary European languages.

What distinguishes Ancient Greek *auto-* from the $_{CC}$markers currently present in the languages of Europe is its remarkable productivity with bases belonging to the class of OBJECTS, STATES or PROPERTIES, with the meaning of 'in itself, *per se*, essentially': *autó·phos* 'the light itself', *auto·thermótēs* 'heat in itself, natural heat', *autó·thermos* 'hot in itself' – compare in Italian *auto·fluorescente* 'fluorescent in itself'. This function appears to be absent or only sporadically attested in today's European languages.

As far as the domain of (both concrete and abstract) OBJECTS is concerned, *auto-* also shows – as already observed for $_{CC}$markers such as German *eigen-* or Finnish *oma-* – the function of contrastive possession (personal belonging, peculiarity): *autó·kheir* 'by one's own hand', *auto·nómos* 'which is governed by own laws, lit. of own law'; moreover, there are also examples with states: *autó·boulos* 'acting by his own will, lit. of own will' (compare to German *eigen·willig*), *aut·eksoúsios* 'who has absolute power' (comparable to *eigen·mächtig*: *eksousía* 'faculty, license, freedom' from *ékseimi* 'to be lawful, granted, possible; it is possible'). In this function, *auto-* used to be in competition with another marker, *idio-* – *idio·genes* 'of its own, particular nature' – which is also attested with bases of the class of ACTIONS – *idio·sunkrasía* 'particular temperament, lit. own, particular mixture' from *sunkeránnumi* 'to mix together'.

The analysis carried out on the $_{CC}$markers of Ancient Greek shows how the situation of this language – whose attestations extend over a very long period of time (about a thousand years in the sources of ROCCI) and include a high variety of text types – is only partially comparable to that of modern European languages. If we consider in particular the languages that today display an outcome of Ancient Greek *auto-* as a productive $_{CC}$marker – i.e. Modern Greek, Romance languages and Maltese – it can be noted that productivity is largely ensured by formations with bases belonging to the class of ACTIONS. However, perhaps because of the influence of *katarévousa* on *dimotikí*, *auto-* in Modern Greek is also present in some formations with bases of the class of OBJECTS (see Chapter 5, section 5.2.3. above).

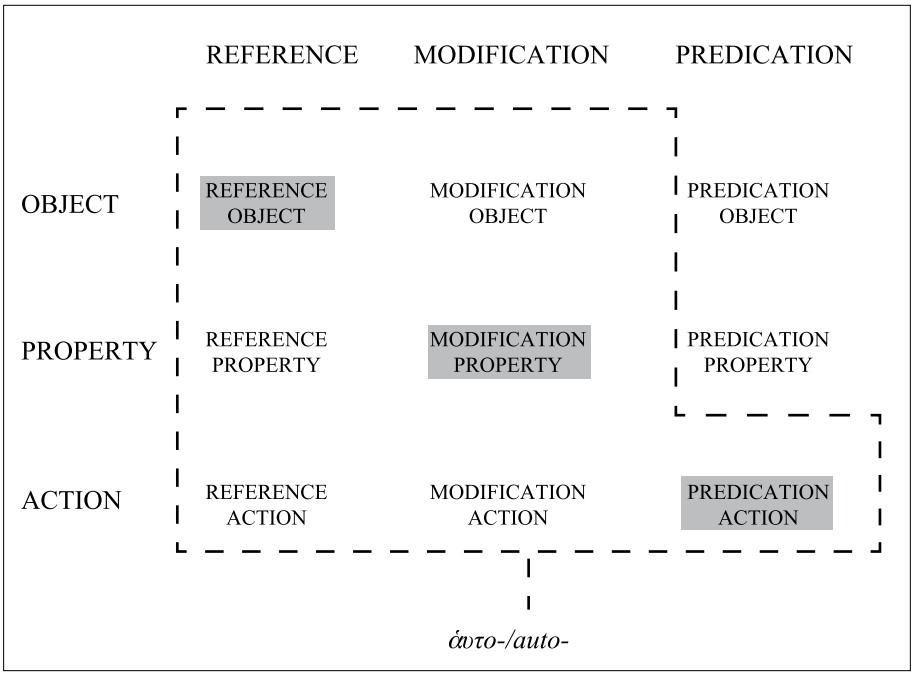


Figure 26. Semantic Map (4): Ancient Greek *auto-* and its compatibility with POSs

We have already considered this discrepancy between Modern Greek, Romance and Maltese languages in Chapter 5. In the next few sections, we will consider the possible influences that most likely produced the observed discrepancies in the data of a variety of languages, and we will propose a diachronic explanation for them.

6.2.2 Latin

Before examining the first available attestations of $_{CC}$markers in Germanic and Slavic languages, let us briefly consider the situation of Latin.

In this case, instead of starting from translation correspondents of formations in languages already analyzed, we proceeded on the basis of the conclusions obtained from the analysis of modern European languages: we have thus browsed the lexicon in search of complex lexemes characterized by the presence of the typical sources of $_{CC}$markers. In Latin, as mentioned in the previous paragraph, the intensifier – that is, the functional category to which most of the sources of $_{CC}$markers belong – does not appear in any $_{CC}$formation. Other sources that are

frequently used as $_{CC}$markers in word-formation are the reflexives and – in the case of markers with a function of contrastive possession – attributive intensifiers. Concerning the latter, in Latin this function was performed by the intensifier itself, i.e. *ipse* (Puddu 2005: 200), while the adjective *proprius*, *-a*, *-um* – from which the attributive intensifiers of many Romance languages will emerge – instead has the lexical meaning 'peculiar, characteristic' (similarly to Ancient Greek *ídios*, *-a*, *-on*): this adjective does not form compounds in Classical Latin either.[146]

A more articulated discussion is needed if we consider the reflexive pronoun *se*. A handful of formations present in today's Italian, such as *seducente*, *semovente*, *sedicente,* may suggest the existence in Latin of some kind of productive process involving the reflexive pronoun *se*.[147] If we exclude the adjective *sedicente*, based on French *soi-disant* (1470, TLF), the origin of the other two adjectives cited above can in fact be traced back to two verbs already attested in Classical Latin: *sēmŏvĕo* 'to remove, to move away; set aside' and *sēdūco* 'lead, draw aside; seduce, corrupt'. It should be emphasized that the present participle *semovente* (= 'moving on its own') implies an interpretation of the element *se-* different from that of the Latin verb – '(to move) by itself' vs. '(move) away' – and must be considered a later construction obtained by reanalyzing the adjective as *se* (refl.) *movente* (LTL). As a matter of fact, in Classical Latin the prefix *sē(d)-* has a different value, conveying the idea of deprivation or removal, and is attested in a large series of more or less idiomatized lexemes (e.g. *sēcūrūs* < *sē* (≈ *sĭne*) + *cūra*), as well as, in isolation, as a privative preposition whose meaning is similar to *sĭne* (+ ABL) 'without' (WALDE-HOFMANN *sub voce*). In verbs, the prefix indicates separation, but has the value of 'aside, apart', therefore of a separation from a whole, and at the same time of isolation from that whole. Think of the difference between *exclūdo* 'do not let in, keep out; remove' and *sēclūdo* 'close outside, isolate; to separate'.[148] In WALDE-HOFMANN (506-507), *sē(d)* is, however, etymologically

[146] Formations with *proprio-* in Italian, as *proprio·cettore* (first attested in 1958), *proprio·cezione* (20th century) (GRADIT) are in fact contemporary technical-scientific terms borrowed from English – *proprio·ceptor*, *proprio·ception* (1906, OED).

[147] The presence of *suicidio* in Italian and of similar lexemes in Romance languages is actually a loan from French, created by using Latin neoclassical material (DELI).

[148] Other verbs with this structure identified in CALONGHI are: *sēcedo* 'to move away', *sēcerno* 'to secern, to separate', *sēcubo* 'to sleep alone', *sēgredior* 'to go aside', *sēgrego* 'to separate, divide from the flock', *sēiugo* 'separate', *sēiungo* 'separate, divide, disjoin', *sēligo* (*se(d)-* + *lego*) 'choose', *sēparo* 'separate, disjoin; distinguish', *sēpono* 'put away, set aside, keep', *sēvehor* 'move away from', *sēvoco* 'call aside, apart, recall'.

connected to the root of the reflexive **sụe*/**se*: from an original meaning 'for oneself', it would then develop on the one hand the value 'isolated, without', and on the other the value 'aside'.

The step of reanalysis undergone by the prefix *se-* within the lexicon, from a prefix conveying the idea of distancing to reflexive, is probably mediated by the verb *sēdūco*, which from the original meaning of 'lead/draw aside', develops in Late Latin in the writings of Christian authors the value of 'corrupt, deceive', thus 'seduce'. In this value, it is possible that a reanalysis (consider *dūco* 'lead, conduct') of the verb as 'lead with oneself, to oneself' took place later. It remains to underline that this reanalysis never led to the development of a truly productive procedure either in Latin or in Romance languages. It is on the other hand questionable whether in Latin the prefix *se-* has ever developed the semantic shift from locative to reflexive.

We can therefore conclude that in Latin no process of word-formation comparable to those seen in the last chapter for contemporary European languages ever existed. The only procedure with native material that could at best correspond to a procedure with $_{CC}$markers has an affixal character, and, as in Breton (see 6.1.2 above), emerges from the verbal domain by an element that acts as a prefix. It is, however, doubtful that the prefixation with *se(d)-* had in (even late) Latin the distinctive semantic characteristics of $_{CC}$markers.

A final addition to complete the survey of Latin concerns *auto-*. In Latin, some formations with *auto-* are attested, but they are very few,[149] always formed using exclusively Greek elements and mostly having a direct correspondent in Ancient Greek. It can be safely concluded that they are therefore simple loans from Ancient Greek.

[149] The formations with *auto-* attested in Latin according to LTL are: *authēmĕron* (*authḗmeron* 'medicina vel pharmaca quotidiana'), *authenta* (A. Greek *authéntēs*), *authenticus* (A. Greek *autentikós*), *authepsa* (A. Greek *authépsēs*), *autochthŏnes* (A. Greek *autókhthones*), *autŏgĕnis* (*autogenḗs*), *autŏgrăphus* (A. Greek *autógraphos*), *autŏmătŏpœētus* (A. Greek *automatopoiētós* 'automatorum ad instar factus, qui per se ipse movetur'), *autŏmătum* or *autŏmăton* (A. Greek *autómaton*), and some derivatives thereof, *aūtŏpractor* (*autopraktōr* 'qui aliquid per se gignit aut perficit'), *autŏpȳros* or *autŏpirus* (from A. Greek *autós* and *purós* 'wheat': 'panis cibarius, hoc est factus ex farina, cui nihil, aut sane parum admodum furfuris demptum sit', therefore 'pure wheat, wheat in itself').

6.2.3 Contrastive coreference markers from native elements

In the previous paragraph, we have seen how in Latin there is a substantial lack of word-formation processes comparable to the $_{CC}$markers of contemporary European languages. Proceeding chronologically from Antiquity towards contemporary Europe, the first European languages for which substantial evidence is available are the Germanic languages and among them Gothic.

It is a well-known fact that the attestations of Gothic consist for the most part of translations of the Bible from Greek, carried out in Moesia by the Arian bishop Wulfila in the second half of the fourth century. However, only a few parts of the translation have been conserved down to the present time: the Gospels and the Letters of Paul from the New Testament, and some fragments of the book of Nehemiah from the Old Testament (Dolcetti Corazza 1997).

As for the formations with $_{CC}$markers, only two compounds are attested containing the compound unit *silba*-, whose corresponding syntactically free element is the intensifier *silba*. The two formations in question are *silba·siuneis*, New Testament Greek *autóptēs*, 'eyewitness' in Luke 1,2,[150] and *silba·wiljis*, Greek *authaíretos*, 'volunteer' in Corinthians II 8,3. Both formations would appear to be structural calques of Greek formations involving both the left elements *autós* and *silba*, and the right elements. We observe a perfect correspondence in the case of *silba·siuneis*, where *ópsis* has the same meaning of *siuns*, i.e. 'view'. The correspondence is lower in the case of *silba·wiljis*: the head of *authaíretos* is in fact a deverbal adjective *hairetós* 'chosen' from the Greek verb *hairéō* meaning 'to take, to grab; to have in one's own power; to choose', while *wilja* 'will' is a noun. The relationship between the constituents and the general meaning of the compound is also the same in the two languages, especially for *silba·siuneis* and *autóptēs*, both paraphrasable as 'he saw with his own sight'. It is possible that in the case of *silba·wiljis,* the presence of a translation correspondent that does not mirror completely the form of the New Testament Greek compound may depend on the fact that to render the sense of 'voluntary', a calque using a verb with a

[150] In the Gothic and Greek texts *silbasiunjos* and *autóptai* respectively, i.e. plural nominatives. The singular nominative is the one proposed by Streitberg (2000) in the dictionary attached to the Gothic Bible edited by Streitberg himself. Dolcetti Corazza (1997: 31) nonetheless proposes a nominative singular form *silba·siuns* of the class in *-i*, instead of the class in *-ja* of the form *silba·siuneis*, in accordance with the inflectional class of the noun *siuns* 'view'.

meaning similar to *hairéō* would not have been transparent, or because the compound *silba·wiljis* was already available in the Gothic lexicon and suitable for translating *authaíretos*.

Not all compounds with *auto-* of the original New Testament Greek, however, have been translated in Gothic using compounds with *silba-*. There are a total of 14 occurrences of *auto-* in the New Testament, of which 6 correspond to passages whose Gothic translation is preserved. Apart from the case of *authaíretos*, which in another passage is rendered by *silba wiljands* – written as a phrase in Streitberg's edition, but also interpretable as a compound[151] – other occurrences of the New Testament Greek *auto-* are translated in Gothic with lexemes of equal meaning, but not by compounds.[152]

It can be concluded that the two formations with *silba-* found in Gothic undergo, like much of the lexicon found in the Gothic Bible, the influence of the New Testament Greek source. It should be nevertheless noted that the form *silbawiljis* is only partially justified by a comparison with the Greek source.

While for Gothic a direct influence of Biblical Greek is obvious, for the ancient phases of English and German the problem arises of verifying whether the model of the outcomes of **selba* as $_{CC}$markers is dependent on models originating in other languages – such as Greek. In the ancient phases of these languages, we attest formations with a native $_{CC}$marker (with *self-* and *selb-* respectively), too. Some of them – *self·wille* and *selb·willīg* – are comparable to the Gothic compound *silba·wiljis*.[153] However, the classical language that could have acted as a model for English and German in their ancient phases – which conventionally last up to the 12th and 11th centuries, respectively – is rather Latin than Greek. As we have seen, however, Latin does not display any $_{CC}$marker. As for Greek compounds with *auto-* in the New Testament, the text of the Vulgate of St. Jerome – which is

[151] See also in modern German for the cases of use of the participle present with *selbst-*, in which the spelling can be as a single word (*selbstbewegend...*) or as separated words (*(sich) selbst bewegend...*).

[152] The verb *authentéo* 'I rule absolutely, I have full authority' is rendered using the verb *fraujinon* 'dominate', derived from *frauja* 'lord'; *autarkeía* in two different occurrences is coherently translated with *ganauha* 'sufficient (n.), sufficient means' from *ganohs* 'quite, a lot'; *autómatos* '(which moves) by itself, (which moves) spontaneously' is rendered simply using the intensifier *silba*.

[153] «In Old English the number of recorded compounds is 13, of which half exhibit the prefix in the objective relation. The only survivals of the Old English compounds in Middle English are *self-will* n. and its cognates; these, together with the plant-name *self-heal* n. (which may also have been common Germanic) are the only representatives in that period of the prefix-formation.» (OED) We will examine examples like *self-heal* later in this paragraph.

the most widespread biblical text in Latin circulating in medieval Europe – never presents morphologically complex formations that could have constituted a model for the formations of German and Old English.[154]

In the historical lexicographic sources, we can also find some formations attested in the ancient phases of English and German concerning a domain of the lexicon, such as that of popular botanical terms, which easily hosts native lexical structures. We found two formations that have correspondents in German and English, are structurally similar and date back to the higher stages of both languages:

i. Old High German *selb·ezza* («Kreuzkraut [...] lat. senecio» KÖBLER) and Old English *self-æte* ('wild oat?' BOSWORTH-TOLLER)
ii. Old High German *selb·heil* («nhd. Weisenaugentrost, [...] lat. euphrasia» KÖBLER) and Old English *self-heal* (prior to 1400 «A name for various plants believed to have great healing properties, esp. Prunella vulgaris (Common S.), Sanicula europæa, and formerly Pimpinella Saxifraga.» OED).

If we turn to Slavic languages and especially to Old Church Slavonic, we have a situation similar to that of Gothic, but about six centuries later. The first attestations in Old Church Slavonic date back to the tenth century, and are considered to be the result of the evangelization promoted by the Byzantines, which affected the South Slavic area starting in the second half of the ninth century with the mission of Constantine (later called Cyril) and Methodius in Moravia (Central Europe, in an area corresponding to the current eastern Czech Republic and western Slovakia).[155] The affinity with the case of Gothic lies in the fact that the evangelization of the Slavs proceeded, at least initially, from the Byzantine Christian East rather than from Latin Christianity, with the consequence that in shaping the first literary attestations in a Slavic language, Greek rather than Latin was the model.

[154] To give an example, in the passages in which the Greek text has *authaíretos* and *autóptēs*, and the Gothic Bible has the formations seen above, the text of the Vulgate has *voluntarii* 'volunteers' and the periphrases *qui ab initio ipsi viderunt* 'who from the beginning saw for themselves', respectively.

[155] Among the Glagolitic codices, on average older than the Cyrillic ones, the Kiev folios (dated to the 10th century) are recognized as a Moravian redaction. This does not necessarily mean that the archetype was composed in Moravia, but it simply means that the peculiar linguistic form of the code is recognizable as belonging to the Moravian area and the local uses of scribes and *scriptoria*. On the problems of interpretation concerning Old Church Slavic see Picchio (1980).

Therefore, if we examine the attestations of formations with *sam-(o-)* in Old Church Slavonic, in many cases we can go back to a Greek formation that constitutes the model. Among the passages of the New Testament in which there are formations with *auto-*, only two are preserved in the Old Church Slavonic translation: of these, only one is a formation with *sam-(o-)*. In Table 17, we can see the other formations attested in Old Church Slavonic, and for how many of them a New Testament Greek model is recognized.[156]

Table 17. $_{CC}$Formations with *sam-(o-)* in Old Church Slavonic (Kurz)

lemma in Kurz	**correspondent in New Testament Greek**	**Greek translation**
presumable borrowings		
samovidьcь 'eyewitness'	*autóptēs, epoptēs*	idem
samovlastьnъ 'autocratic, high-handed [eigenmächtig]'	*auteksoúsios*	idem
samovlastьcь 'monarch'	–	*aútarkhos*
samovolьnъ 'voluntary'	*authaíretos / hekoúsios*	idem
samoglasьnъ (liturgical chant)	–	*idiómelos* / *autómelos*
samodrъžitelь 'monarch'	*autokrátōr*	idem
samodrъžьcь 'monarch'	*autokrátōr / tropaioïkhos*	idem
samodělьnikъ 'author, creator'	*autourgós*	idem
samoprichodęi 'defector	*ho automolõn*	idem
samotvorьnъ 'earned by oneself'	*idióktētos*	idem
presumably independent formations		
samoizvolьnъ 'voluntary'	*ethelóthuton*	idem
samoljubьcь 'selbstsüchtiger Mensch'	*phílautos*	idem
samorastyi 'that arises from itself, rural'	*automátōs phuómenos*	idem
samochotьně 'on one's own will'	*hekőn*	idem
other formations		
samovidьně 'evident'	–	–
samovlastь 'free will'	–	–
samochotь 'free will'	–	–

[156] The formations with $_{CC}$marker in Old Church Slavonic are taken from Kurz.

From the table we can see that in many cases (12 out of 17) a Greek archetype has been recognized (see "**correspondent in New Testament Greek**" in Table 17), only in nine cases a compound with *auto-* could have constituted the model (see "***presumable borrowings***" in Table 17). In the remaining six cases, in one case (*samorastyi*) there is an adverbial phrase plus participle, where the adverb is a compound with *auto-* (*automátōs phuómenos*), and finally, in one case (*samotvorьnъ*) we have a compound with *idio-* (*idióktētos*). In two cases (*samovlastьcь* and *samoglasьnъ*), a Greek archetype cannot be identified in the New Testament, but compounds with *auto-* or *idio-* as possible correspondents or sources of the Slavic formations can still be proposed (*aútarkhos* and *idiómelos* / *autómelos*, respectively).

Consequently, Old Church Slavonic formations with *samo-* show that they depend on (New Testament) Greek lexical models even more than Gothic. At the same time, many formations are structurally rather independent from the Greek model, which they translate to the point that it can be concluded that their formation is independent from and perhaps chronologically preceeding the Greek models. The case of *samorastyi* is significant because it structures the concept which in Greek is expressed through an adjectival phrase (*automátōs phuómenos* 'spontaneously born') with a compound. But even more interesting is *samochotьně*, which, despite having a structure similar to Greek *autokínētós* or *automátōs*, therefore, literally, 'that moves by itself', translates instead the Greek adverb *hekṓn* 'voluntary, spontaneous'.

Summing up, in the ancient phases of some European languages (Gothic, Old Church Slavonic) the direct influence of Greek can be recognized. One could argue that Ancient Greek constitutes the source model of $_{CC}$markers in Germanic and Slavic languages, where an autochthonous model was simply absent before Greek influence. However, if on the one hand we consider the discrepancies with respect to the model language that can be observed in the formations attested in sources translated from Greek and, on the other hand, the fact that languages such as German and English – which in the ancient phases did not undergo a Greek influence – display arguably indigenous formations, we can maintain that $_{CC}$markers were already present in Germanic and Slavic languages.

However, a different picture must be drawn regarding the subsequent productivity of these processes. If we take English as an example, in the lemma relating to the so-called prefix *self-*, the editors of the Oxford English Dictionary observe:

(145) *Self-* first appears as a living formative element about the middle of the 16th cent., probably to a great extent by imitation or reminiscence of Greek compounds in ἀυτο-. The number of self-compounds was greatly augmented towards the middle of the 17th cent., when many new words appeared in theological and philosophical writing, some of which had apparently a restricted currency of about 50 years (e.g. 1645–1690), while a large proportion became established and have a continuous history down to the present time. (OED-online)

In Figure 27, we can observe the increase of $_{CC}$formations with *self-* in English across the centuries. By dividing the formations with *self-* according to the date of first attestation in periods of 50 years, an evident growth in the number of formations is observable only during the seventeenth century, therefore in conjunction with the development of modern science, but also precisely with the progressive diffusion of the knowledge of Greek in Europe (see below in 6.2.4). Another peak, however, occurs between the early 1800s and the first half of the 1900s – the same period in which, as we will see in 6.2.4 as well, *AUTO-* emerges as a productive element in Romance languages.

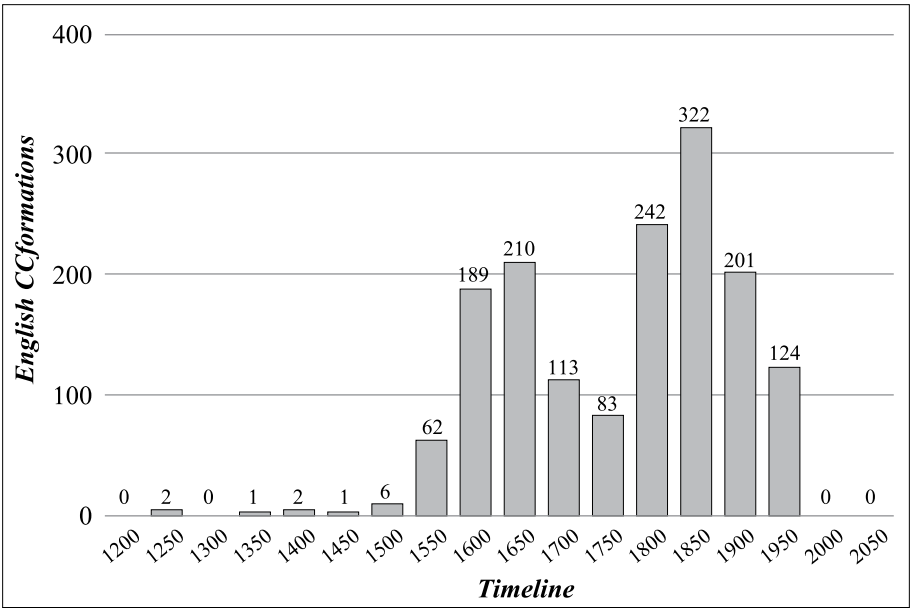


Figure 27. Increase of $_{CC}$formations with *self-* in English (source: OED-online)

In addition to the influence of Greek, the latinate stratum of the English lexicon should also be considered. This stratum appears to be compatible with the process considered, but also, since the 16th century many of the formations with *self-* have had as their basis lexemes belonging to this non-native layer: in the formations attested between 1550 and 1599, the formations with a [+ latinate] basis are 21 out of 62 (33.87%), while in the following fifty years they grow to 130 out of 189 (68.8%).

A last noteworthy phenomenon is the loss of compatibility of *self-* with bases belonging to the class of OBJECTS, and the concomitant loss of the value of contrastive possession. The clues pointing to this development consist in the fact that the OED indicates as obsolete the formations attested between the beginning of the 1600s and the middle of the 1700s, in which *self-* indicates «relating to oneself, one's own, personal, individual, private, intimate». Consider the example in (146) below:

(146) ***Selfe-guiltinesse*** *commonly makes men partiall, in judging others.* (1641, D. Cawdrey)

In the example *selfe-guiltinesse* means 'one's own guilt' or 'personal guilt' and thus indicates a contrast between one's own responsibility in a deed and the responsibility of someone else.

In a slightly earlier period (the last attestations of the OED date back to 1654), we also observe the disappearance of the use of *self-* accompanied by a possessive as an attributive intensifier as in (147).

(147) *He considered that many Towns in Picardy were under the* ***self*** *Forces of Spain* (1654, Earl of Monmouth)

In this case, the use of *self* basically corresponds to that of *own* ('under Spain's own forces'), and in fact, from that moment on, the function of the attributive intensifier in English will be exclusively performed by the adjective *own*, whose first attestation in the OED as an attributive intensifier, reproduced in (148), is however recorded already in the mid-16th century:

(148) *If they would keep their* ***own*** *counsel, he, for his part, would never confess any thing to die for it* (1551, PF Tytler)

6.2.4 Romance languages and neoclassical lexicon

We have already noted in 6.2.2 above that in Latin no $_{CC}$marker is available, if we exclude the cases of Late Latin formations with the prefix *se(d)-* and of the few borrowings with *auto-* from Greek. A similar situation characterizes Romance languages up to the second half of the 19th century. Only in recent times did Ancient Greek *auto-* start to occur in these languages in an increasingly rich series of formations, until it became a fully productive process during the 20th century. Let us retrace the stages of this development in more detail.

In the ancient phases of Romance languages, Latin was the main language of culture, so a small number of Greek borrowings were recorded in the lexicon. This situation has a turning point with Humanism and with

«the massive adoption (especially when compared with the small group of Greek loanwords of Middle Latin tradition, many of which linked to the ecclesiastical or rhetorical sphere, already present in the ancient vernaculars) in the various philosophical, historical, rhetorical-poetic and scientific sectors of Classical Greek lexical forms which did not make their way not only in Latin and Medieval vernaculars, but sometimes not even in Classical Latin.» (Tesi 1994: 28)[157]

Before 1400, only a few formations with *auto-*, more precisely *authentic* (1304-1308 GRADIT and DELI) – French *authentique* (1211 REY) – the corresponding adverb *autenticamente* (1357) and its derivative *autenticare* (1400) are attested in the Romance vernaculars – but they were already attested in Medieval Latin. The absence of the linking vowel *-o-* in the Greek compound *authéntes,* due to the presence of a rough breathing on the second element, made the formation unanalyzable.[158]

As for the knowledge of Greek, it should be noted that the first chair of Greek in Western Europe has been endowed in Florence in 1397, and that before this

[157] «[L]a massiccia riassunzione (specie se confrontata con l'esigua schiera di grecismi di trafila mediolatina, molti di ambito ecclesiastico o retorico, presente nei volgari antichi) nei vari settori filosofici, storici, retorico-poetici e scientifici di forme lessicali greco-classiche non penetrate non solo nel latino e nei volgari medievali, ma talora neppure nel latino classico.» (Tesi 1994: 28; my translation)

[158] According to Beekes (2010: 169), *authéntes* «point[s] to earlier *ἔντης, the full grade of the root of ἀνύω 'to accomplish'», a fact that likely rendered the compound opaque already in the Classical Greek period. More recently, the opacity of loanwords such as *authentique* is also demonstrated by the presence of orthographic variants such as *auctentique* (1403 REY) in French, probably due to the influence of the outcomes of Latin *auctoritāte* (m).

date, the knowledge of Greek and its literature was most likely limited to very few people. The scarce diffusion of the knowledge of Greek must have remained so until the end of the fifteenth century – that is, until after the fall of Constantinople (1453) and the subsequent movement of Greek-speaking exiles from the areas conquered by the Turks. On the other hand, the fact that many Greek texts circulated primarily in Latin translation even after the rediscovery of Greek in the West, made humanistic Latin the true and major vector of Greek loanwords into the lexicon of European languages (Tesi 1994: 37-41, n. 1-3). While vernaculars gradually gain importance and domains of use at the expenses of Latin starting with the Humanism, Latin and Greek continue to constitute a stock of lexical material for technical-scientific terminologies (Marello 1996: 53-58). The formations with *AUTO-* drawn directly from Greek date back to this period. From the first attestations indicated by GRADIT and REY, it appears that the Italian formations are attested later and often presumably introduced through French. The formations recorded for Spanish in COROMINAS are generally the same ones that are attested in the same period in French, but, similarly to Italian, they tend to be attested later (see Table 18).

TABLE 18. Dates of first attestation of Ancient Greek loans with AUTO- in Italian, French and Spanish

Italian GRADIT	**French** REY	**Spanish** COROMINAS
autarchia (1829)	*autarcie* (1793)	*autarquía* (1884)
autocrate (1788)	*autocrate* (1768)	*autócrata* (1835)
autocratore (1793)		
autocrazia (1765)	*autocratie* (1794)	
autoctono (1796)	*autochtone* (1560)	*autóctono* (1884)
	autodidacte (1557)	
autografo n. (1745)	*autographe* n. (1580)	
automa (1589)	*automate* (1532)	*automato* (1582)
autonomia (1739)	*autonomie* (1596)	*autonomia* (1702)
autoschediasma (1728)		

French, in addition to having a higher number of attestations, is credited as a source of some Greek borrowings in Italian, as *autocrate* from French *autocrate* and *autografo* fromFrench *autographe*. These examples are a sign of the cultural

influence of French on Italian, which corresponds to the increase of the political influence of France on Italy, whose peak will occur in the 18th century and in the Napoleonic age. The cultural influence, due to the knowledge of the language and the spread of French and French publications, continued in Italy between the 18th through the first half of the 20th century (Marello 1996: 64-66). But in this period, French was important not only in Italy, and its political and cultural centrality, before giving way to the Anglo-American socio-cultural hegemony, has left important traces in technical and scientific lexicon. In this time, a number of neoclassical formations, unknown to classical Greek, were produced from Greek material. Furthermore, a new wave of formations emerged in this period, combining a Greek element with a Latin one.

Concerning the formations displaying *AUTO-*, we have examples of both kinds. Formations unknown to Ancient Greek are, for example, the French formations in (149):

(149) *autophage* (1854 REY; first attestation in Italian 1875 GRADIT)
autoscopie (1903 according to REY, but OED attests it in French already in 1860)
autotomie (1882 REY)

All these formations are subsequently borrowed into English and German, as can be seen from their first attestations in (150).[159]

(150) English: *autophagy* 1860, *autoscopy* 1861, *autotomy* 1887 according to OED
German: *Autophagie*, *Autoscopie* 1863, *Autotomie* 1884

Examples of Greek-Latin hybrid formations with *AUTO-* are the ones in (151), which are also later attested in English and German.

(151) French: *autoclave* (1820 REY; first attestation in Italian 1865 GRADIT), *automobile* (1861 as adjective REY; first attestation in Italian 1892 GRADIT)
English: *autoclave* 1820, *automobile* 1876 as adjective, 1881 as noun
German: *Autoklav* 1822, *Automobil* 1896

[159] The dates of the German forms, when indicated, are those reported in the corresponding English lemmas of the OED.

The stream of influences across European languages is therefore multi-directional and mutual, and if there has been an enduring importance of French in producing and spreading neoclassical formations, especially in the technical-scientific domain, other culturally and politically hegemonic European languages, such as German and English, have quickly adopted many French formations with *auto-* and have contributed to their diffusion. The most surprising consequence of the spread of these words and concepts, however, concerns Modern Greek. In this language, both formations with *afto-* dating back to Ancient Greek (*aftókhir* 'suicide', from the aforementioned Ancient Greek *autókheir*) and formations based on European models (*aftokínito* 'automobile' calqued on French *automobile*, *aftokatastrofí* 'self-destruction', first attested in 1878 according to elM1, calqued on English *self-destruction*) occur side by side. It is worth noting that in Modern Greek, French Greek-Latin hybrids are turned into calques made up only of Greek lexical material.

Turning back to Italian, the importance of French has been increasingly undermined, especially during the twentieth century, by the growing importance of English borrowings. Some of them are semantic calques or translations – often of the latinate stratum, such as *coalizione* from *coalition* or *conformista* from *conformist* – adapted loans – *assenteismo* English *absentee·ism*, *rostbif* (an obsolete orthographic form of) English *roast-beef* – and finally, more recently non-adapted loans. Focusing on translations, a particular type of them is the structural calque, in which a complex formation in one language is structured with native material in another language, while maintaining the motivational relationship (at least at the lexeme level) between the elements involved. Think of the case of the Italian VN compound *grattacielo* to reproduce the English synthetic compound *skyscraper*.

At the beginning of this section, we have anticipated that in recent times, Romance languages have experienced an increase in the number of formations with *AUTO-* connected to the crucial innovation of applying the process – which previously involved exclusively learned bases of the Greek stratum or to a lesser extent of the Latin stratum – to Romance native bases.

In Italian, it is possible to observe the presence of formations with *auto-* dating back to before the nineteenth century with exclusively Greek material, while the Greek-Latin hybridization occurs later, and follows, as we have seen, French models. On the other hand, the hybridization between *auto-* and Italian bases occurs only as late as the end of the 19th century. Iacobini (2003) proposed that Eng-

lish influence triggered the productivity of Italian formations with *auto-*, without mentioning a possible mediation from French. Certainly, the formations indicated by Iacobini as an example (e.g. *self-government*, *self-defense*) are attested earlier in English than in any other European language, since they date between the 16th and 17th centuries. Furthermore, the distance between the first Italian and French attestations of these formations does not allow any speculation concerning the direction of a possible influence (see Table 19).

Table 19. Interaction between English self- and Romance AUTO- in French and Italian

English OED	**French** TLF	**Italian** GRADIT
self-defence (1651)	*auto-défense* (inizio XX sec.)	*autodifesa* (1896)
self-government (1691)	(*autogouvernement*)	*autogoverno* (1890)
self-moving agg. (1607)	*automobile* agg. (1866)	*automobile* agg. (1892)
self-accusation (1662)	*autoaccusation* (1903)	*autoaccusa* (1960)
self-criticism (1857)	*auto-critique* (1866)	*autocritica* (1892)

However, it should be remembered that at the turn of the twentieth century, French is still a crucial contributor of neoclassical technical terms that radiate towards Italy, but also towards the British Isles – consider lexemes such as *photography*, *bureaucracy*, in addition to the aforementioned *automobile*. It is therefore probable that, even if the models for single formations with *auto-* indeed came from English – which they certainly did with increasing frequency and without intermediaries starting from the mid-twentieth century – the neoclassical pattern with the combining form *auto-* is probably drawn from the Romance language with the highest degree of innovations and productions in terms of neoclassical coinages and Greek-Latin hybrids, i.e. French.

Obviously, the routes of contact and of lexical borrowings are anything but univocal and unidirectional. An example is the presence in English of some formations with *auto-* at the end of the nineteenth century, along with a second element of the latinate stratum which might suggest a French influence on English, although the dates of first attestation do not support such conclusion – English *auto-destruction* dates back to 1888 (OED), while French *auto-destruction* to 1898 (TLF).[160] A second example is a lexeme such as Italian *autocoscienza*, which is

[160] It must be noted, however, that *self-destruction* is attested prior to 1586, according to the OED.

recognized as a calque on the German *Selbst·bewusstsein*, rather than on English *self-consciousness* (Klajn 1972: 127).

The last step of the path from Ancient Greek to modern languages therefore takes place after the second half of the 19th century, with the joint action of the flourishing neoclassical lexicon and the model of English formations with *self-*.

On the one hand, the presence in the neoclassical lexicon of formations with a stressed second element which can also occur as a syntactically free noun (e.g. Italian *(auto)gráfico*, *(auto)biografía*, *(auto)móbile*), makes these formations more easily segmentable in comparison, for example, with cases of tonic accent on the first element as in *autóctono*, or at the morphological boundary like *autóptico*, or the cases of crasis of the vowel *-o-*, as in *autarchía*. Segmentability probably favoured the diffusion of the process beyond specialized fields. Furthermore, the English models, although displaying the native element *self-* as their first element, have in many cases a latinate base (see the examples in Table 19), which has certainly facilitated their segmentability and therefore their borrowing into Romance languages.

A final observation on the complex pattern of borrowing of Greek *auto-* into Romance languages that shows the mediation of English concerns the functions of these $_{CC}$markers and their pattern of compatibility with the bases with which they combine. Due to the emergence of the attributive intensifier *own,* the progressive exclusion of *self* from attributive intensification has been observed in English. In word-formation, this exclusion corresponds to the progressive exclusion of *self-* – but the same holds for German *selbst-* – from the compatibility with bases belonging to the class of OBJECTS and, more generally, from the function of contrastive possession, which is instead conveyed by *eigen-* in German, and expressed with the adjective *own* in English.

On the other hand, we have noted that *afto-* in Modern Greek and especially *auto-* in Ancient Greek are compatible with bases belonging to the class of OBJECTS and with the function of contrastive possession. The mediation of the more recent English model can explain the discrepancy that we have observed in Chapter 5, section 5.2.3, and above in 6.2.1, between *AUTO-* in Romance languages and the $_{CC}$markers of the ancient and modern phases of Greek, a discrepancy that would not be otherwise easy to explain.

6.2.5 Eastern and northern European languages

The path that has led to the development of the prefix *AUTO-* in Romance languages certainly requires a deeper scrutiny of the hypothesis that the Greek influence during the humanistic and Modern Age also affected the development of markers in languages, like English, where an indigenous $_{CC}$marker is available. However, though eventful, that path is nonetheless quite clear.

In this paragraph, we will consider a set of clear lexical parallels between complex lexical items in different East and North European languages. In contrast to the case of *AUTO-,* the contact pattern appears less clear and would require, in order to be exhaustively proven or excluded for sure, more solid evidence than the one that we will provide in this chapter.

The phenomenon in question concerns the presence of a double system of $_{CC}$markers in Germanic, Slavic, Baltic and Finnic languages.[161] But more than the presence of two specialized markers on two distinct areas of the conceptual map of POSs, what is striking is the presence in all these languages of a certain number of similarly structured formations using lexical elements of the same meaning, so much so that the presence of structural calques can be identified.

Some examples of these parallels are given in (152).

(152) 'in person': German *eigen·händig*, Croatian *vlasto·ručni*, Polish *własno·reczny*, Lithuanian *sava·rankiškas*, Finnish *oma·kätinen*
'obstinate': German *eigen·willig*, Croatian *svoje·voljan* 'voluntary, capricious'
'arbitrary': German *eigen·mächtig*, Lithuanian *sava·vališkas*
'stubborn': German *eigen·sinnig*, Russian *svoe·nravnij*
German †*eigen·köpfig*, Croatian *svoje·glav*, Czech *své·hlavý*

It can be noted that these formations display some common features: in different languages, comparable elements contribute to structurally similar formations having similar meaning. First, markers whose syntactically free correspondents have functions that are substantially similar occur as modifiers: *eigen* corresponds to *własny* and *oma*, i.e. attributive intensifiers, and *své*, *svoe* and *savas* have (reflex-

[161] For a synchronic survey of this double system of markers, see Chapter 5, section 5.2.3, and Figure 5.

ive) possessive function, just like *eigen* and its aforementioned correspondents.[162]

The bases with which they combine are also comparable: German *Hand* corresponds to Croatian *ruka*, Polish *ręka*, Lithuanian *ranka* and Finnish *käsi*; German *Kopf* corresponds to Croatian *glav*, Czech *hlav*, etc. Furthermore, concerning the affixes used, German *-ig*, Polish *-ny*, Russian *-nij*/*-nyj*, Lithuanian *-iškas*, and Finnish -V*nen* all productively form adjectives from nouns.

It could be objected that with regard to productive $_{CC}$markers which combine with bases belonging to the class of ACTIONS, series of formations are also identifiable across languages – see, for example, the following series of formations of the same meaning:

(153) English *self-pollination*, Italian *auto·impollinazione*, German *Selbst·bestäubung*, Russian *samo·opilenie*, Modern Greek *afto·epikoníasi*, Albanian *vetë·pjalmim*

Undoubtedly, it is possible and even probable that various models are transferred from one language to the other, but in the case of the $_{CC}$markers in the formations cited in (153), these are productive and widespread processes for which it is therefore in many cases difficult to determine whether they are directly based on models in another language, or if they are indigenous formations based on a productive schema.

As for the examples in (152), what is peculiar is the fact that there is a core of similarly structured formations, which use lexical elements characterized by the same semantic motivation, each of which has a meaning roughly comparable to that of its structural correspondents in other languages. In addition to this aspect – shared with the examples in (153) – the formations in (152) are also characterized by the fact:

i. that the markers used differ from the main $_{CC}$marker (*eigen-* vs. *selbst-* and similarly *oma-* vs. *itse-* or *ise-*, *własno-* or *své-* vs. *sam-(o-)*, *sava-* vs. *savi-*);
ii. that the lexical source is the reflexive possessive or the attributive intensifier and – at least in the case of Baltic and Slavic languages – that this scheme is limited to a few formations;

[162] On the marker *sava-*, its formal opposition to *savi-* and the issue of its identification with the possessive marker *savas* see Chapter 4, section 4.2.5, above.

iii. that the contribution of the marker can be interpreted exclusively in the attributive sense of contrastive possession – we can therefore define them as markers of contrastive possession in the strict sense: $_{CP}$markers;
iv. that the meaning of the resulting formations is compositional to a limited extent, and that the pattern of divergence from compositionality is parallel across different languages.

However, as we have seen in Chapter 5, section 5.2.3, in Germanic languages, in Finnish and Estonian, these markers are also compatible with bases belonging to the class of ACTIONS, in which case they can instead convey contrastive coreference as well.

As for Balto-Finnic and Baltic languages, the recent attestation of these languages (apart from sporadic attestations of nouns and phrases, the texts available date back to the sixteenth century) does not allow us to draw conclusions on their most ancient phases, but a diachronic study of the lexicon, starting from the first attestations, would certainly help shed more light on this phenomenon. It is nevertheless possible to hypothesize a dependence on the influence of the Germanic culture in various periods and starting from remote times (see Dahl and Koptjevskaja-Tamm 2001). This can lead us to exclude on the one hand that Balto-Finnic languages could be a propulsive centre of the model underlying the formations listed above in (152). On the other hand, it is also unlikely that the Indo-European languages of the Baltic group developed the model autonomously.

Given the presence of ancient sources, the hypothesis of the influence of German on Slavic languages can be put to the test more concretely by first comparing the current situation with the data available from older attestations. In Table 20, we can see some examples of the parallels between formations with $_{CP}$marker and German formations of similar structure concerning Slavic languages.[163]

[163] The comparison privileges structural comparability between the formations, while the correspondence of meaning may show some discrepancies, for example, Czech *svéhlavý* and Russian *svoenravnyj* translate German *eigensinnig* (also because the structurally more similar German *eigenköpfig* is today obsolete). On the other hand, *svévolný* and *svoevolnyj* translate both the structurally similar German *eigenwillig* and the somewhat different *eigenmächtig*.

It should be noted that for Polish, in the lexicographic sources consulted, no other formations with $_{CP}$marker were found besides those listed in Table 20, a fact that makes Polish a little marginal compared to other Slavic languages considered, including Russian. However, we can find additional correspondents of some German formations with *eigen-* translated with noun phrases with the attributive intensifier *własny* and the order Noun Adjective: e.g. *korzyść własna* 'self-interest [Eigennutz]'. This possibility is also present in Czech for other formations with *eigen-*: *Eigen·kapital* 'equity capital' is translated with *vlastní kapitál*, *Eigen·tor* 'own goal' corresponds to Czech *vlastní branka*. It should be added that in all the Slavic languages considered, it is also possible to find a formation that has 'will' as its basis, which, however, combines with *sam-(o-)* as first element, instead of one of the $_{CP}$markers seen in Table 20. For example, Croatian and Serbian *samo·voljan* 'arbitrary, capricious', Czech *samo·volní* 'spontaneous', Polish *samo·wolny* 'arbitrary', Russian *samo·wolnyj* 'arbitrary'.

Table 20. $_{CP}$Formations in Slavic languages and their German correspondents

German	base	Croatian	Czech	Polish	Russian
eigenhändig	hand	*vlastoručan* (Serbian *svojeručan*)	*vlast-noruční*	*własnoręc-zny*	*sobstven-norúčnyj*
†eigenköpfig	head	*svojeglav*	*svéhlavý*	–	–
eigenartig	manner, temperament	–	*svérazný*	–	*svoenravnyj, svoeobraznyj*
seinerzeitig	time	*svojevre-men*	–	–	*svoevrémennyj*
eigenwillig	will	*svojevoljan*	*svévolný*	*swawolny*	*svoevólnyj*
eigenmächtig	power	(Serbian *svojevlastan*)	*svémocný*	–	*svoevlástnyj*
eigennützig	advantage, profit	–	–	–	*svoekorysť*

Drawing back our attention to the data of Old Church Slavonic seen above in 6.2.3 in Table 17, we can see how in fact a formation similar to the current Croatian and Serbian *samovoljan*, Czech *samovolní*, Polish *samowolny* and Russian *samowolnyj*, i.e. *samovolьnъ*, was already attested in Old Church Slavonic. If we consider on the other hand Old High German, we have a similar situation. At the time, *eigen-* was not available as a compounding element in the func-

tion of $_{CP}$marker (see *eiganbuoh* 'New Testament, lit. book of the legacy, of the inheritance', based on the noun *eigan* 'property, part of an inheritance').[164] Furthermore, formations constituting the ancestors of *eigenwillig* 'voluntary', *Eigenname* 'proper name', †*eigengewaltig* 'arbitrary'[165] were formed with the $_{CC}$marker *selb*: *selb·willīg*, *selb·namo*, *selb·waltīg* (KÖBLER). The availability of *eigen* as a $_{CP}$marker emerges from the Middle High German period, during which *eigen-willec* or *eigen-nuz* appear, but, for example, *sëlp-(ge)walt* resists unchanged (LEXER).

This piece of evidence might support the hypothesis of a German influence: German introduces an innovation, and this innovation expands into neighbouring linguistic areas, such as the Slavic-speaking area. Additional evidence could come from the presence in the Slavic context (from the data we gathered only in the southern and western varieties) of formations all converging on the schema $_{CP}$marker-'head'-suffix: Czech *své·hlavý*, Croatian *svoje·glav*.[166] The peculiarity of this example is that the corresponding German formation, *eigen·köpfig*, is today obsolete in German (the online edition of the DUDEN does not report it), but it is attested in modern German at least since Luther (GRIMM).

The vanishing of *eigenköpfig* in today's standard German and the presence of similar structures in other languages influenced – we hypothesize – by German, could be an example of 'fase sparita' in Bartoli's (1925) terms: the area outside present-day Germany which has been subject to German cultural influence retains forms that in the original area have gone into disuse due to further innovations.

Furthermore, from a historical point of view, the possibility that German may have influenced these languages (Czech, Serbian and Croatian) is very high, considering that these areas of central and southern Europe have been progressively included into the political sphere of the Austrian-Hungarian Empire, and because of that German language and culture have become of crucial importance, espe-

[164] In Old High German, *eigan* as adjective is synchronically derived by conversion from the past participle of the verb *eigan* 'have, own' (GRIMM). The verb became obsolete in later phases also due to the overlap in the Early New High German period with the verb *eignen* 'possess', derived from *eigen* 'own' (cfr. Frühneuhochdeutsches Wörterbuch: http://fwb-online.de/go/eignen.h1.3v_1646838051; accessed 5th May 2022).

[165] It is attested in Luther: 'sua ipsius auctoritate factus, eigenmächtig' GRIMM.

[166] A formation based on a similar schema but with the main $_{CC}$marker is also present in Hungarian: *ön·fejű* 'stubborn', where *fej* 'head'.

cially in Bohemia, but also in northern Croatia.[167]

However, it should be noted, on the one hand, that many areas of present-day Poland, in addition to the obvious geographical contiguity with the German-speaking area, have been subjected at various times to German-speaking rulers (first in the Late Middle Ages the Teutonic Order, then towards the end of the 18th century the Hohenzollerns and the Habsburgs). On the other hand, it is difficult to argue that Russia could have suffered German cultural influences, at least until the period of Peter the Great and his policy of westernization of Russia. On the side-lines of the discussion concerning Slavic languages, it should also be noted that in Hungarian, a language whose lexicon has certainly undergone a strong cultural and linguistic influence from German, the distinction between a $_{CC}$marker and a $_{CP}$marker is absent. In fact, in Hungarian the bases of the class of OBJECTS and the interpretation of contrastive possession – except for the case of the doublet *ön·kezű*/*saját·kezű* 'on one's own', both involving *kez* 'hand', and the latter based on *saját* 'own' – correspond to the use of the same marker *ön-* that combines with bases of the class of ACTIONS, and conveys the function of contrastive coreference.

In conclusion, the presence of substantial German influences does not completely support the hypothesis proposed so far of a contact pattern causing the spread of an innovation that emerged in the German-speaking area towards the Slavic area (and beyond), because the known scope of German influence on eastern Europe is not consistent with the distribution of the supposed calques discussed above.

At this point, we can put forward another explanation of these parallel series of formations, noting first that the process that exploits the possessive reflexive as a $_{CP}$marker is already attested in Old Church Slavonic. The only two formations attested are *svoje·volьstvьnъ* 'of one's own will' and *svoje·plemenьnikъ* 'compatriot'. The presence of this process in Old Church Slavonic can explain its diffusion in all the Slavic languages considered. The alternative process, with the attributive intensifier as a $_{CP}$marker, evidently emerged later, and in fact, it is marginal in South Slavic and has two different forms in Western and Eastern Slavic: the

[167] It is thought that the first works that raised the question of the Czech language and defended its cultural use, dating back to the second half of the 18th century, were not written in Czech, but in German (Dell'Agata 1972). Furthermore, it is well known that literacy in German was very common in Croatia, especially in northern cities close to present-day Slovenian and Hungarian borders (consider e.g. Kordić 2021 on the use of German in the public life of Osijek until the Second World War).

cognates *własn-(o-)* in Polish and *vlastno-* in Czech, on the one hand, and Russian *sobstvenn-(o-)*, on the other.[168]

A second hypothesis can therefore be proposed. In the ancient phases of Greek, German, English and Old Church Slavonic, the element that acts as the adverbial and adnominal intensifier also fulfilled the function of attributive intensifier. This function has gone lost over time and has been progressively adopted by other lexical elements that previously served as adjectives with the meaning of 'particular' (such as the Greek *idiós*) or 'possessed, possession' (such as German *eigen* and the Polish *własny* and its cognates). The rise of these new elements in syntax could have had the result, which at this point would have the character of an autonomous development, of undermining the scope of use of the intensifier in word-formation as well, with the effect that it therefore progressively lost compatibility with the class of OBJECTS and the contrastive possession value, with only the contrastive coreference function left, along with the compatibility with bases in the class of ACTIONS.

With this second hypothesis, the phenomenon must no longer be interpreted as a case of convergence affecting the lexicon, but instead constitutes an epiphenomenon of a case of convergence at the lexicon-syntax interface. In this context, the case of Hungarian would appear as less marginal, because, although Hungarian too has developed a bipartite system differentiating the adverbial and adnominal intensifier (involving the marker *maga*) from the attributive intensifier (involving the marker *saját*) at an earlier historical stage, the marker *ön* (today external to the intensifier-reflexive continuum) covered both the functions of the reflexive/intensifier and the emphatic possessive.

The parallels between German and Slavic languages can certainly be considered largely the result of an internal development, especially regarding the formations in which the possessive reflexive is exploited as a $_{CP}$marker. However, it remains possible that the formations in which the $_{CP}$marker corresponds to the attributive intensifier – certainly later, as shown by the substitution of *svoe-* with *sobstvenno-* in the *eigenhändig* type in Russian (compare ancient Russian *svoe·ručnyj* attested in the 15th century) – may at least partially have developed based on a model available in German.

[168] A South Slavic language such as Slovenian also has the attributive intensifier type with a formant, *lastn-(o-)*, which is cognate to the attributive intensifier of other Slavic languages: e.g. *lastno·ročen* 'autonomously', with the same structure as German *eigen·händig*'.

6.3 Diachrony and contact in the light of Construction Morphology

After having carried out this diachronic exploration of the possible explanations for the divergences from the generalizations and tendencies put forward in Chapter 4 and 5, and after having followed the spread of some $_{CC}$markers across Europe and across the centuries, we can try to model some of these phenomena by applying to them the principles and devices of Construction Morphology.

In Figure 28, we have applied the principle of schema unification to a multilingual setting, such as that of Malta and of the heavily influenced Maltese lexicon.

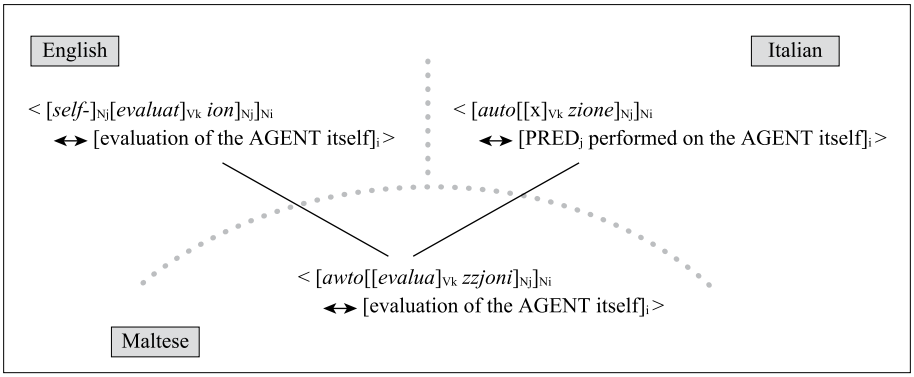


Figure 28. Cross-linguistic schema unification: Maltese *awto-*

The Romance layer of the Maltese lexicon is known to be influenced by Romance varieties, especially Italian and Sicilian. The case of *awto-*, however, calls for a more complex explanation: $_{CC}$formations such as *awto·evaluazzjoni* includes in fact a base that appears derived from English *evaluate*, rather than from Italian *valutare* 'evaluate'. We can thus interpret this as a case of schema unification operating cross-linguistically. The form of the Maltese schema has most likely issued from Italian, but single formations which motivate the Maltese example depend on English models instead.

A similar situation can be that of Italian *auto-*, i.e. a morphological pattern that, as we have seen earlier in this chapter, emerged only in the second half of the 19th century. If on the one hand the French model has been at the core of the formal make-up of early $_{CC}$formations, the later motivation for coining new ones, which is the source of the productivity of the schema, most likely came from English *self*-compounds.

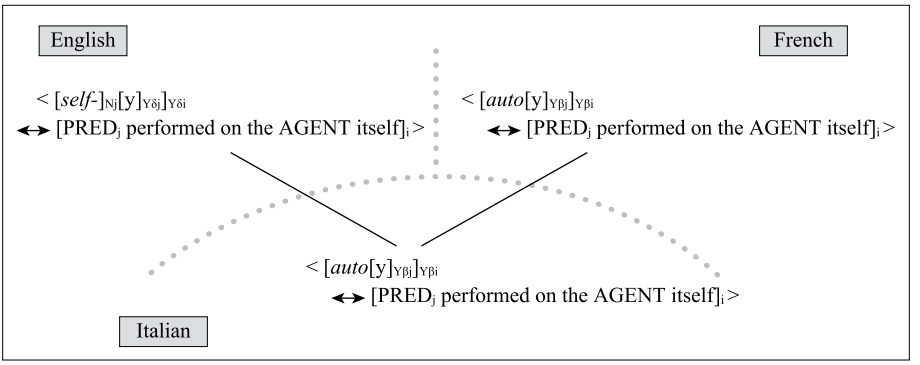


Figure 29. Cross-linguistic schema unification: Italian *awto-*

Despite that, the French model kept on influencing the Italian, which is on the one hand compatible with verbal bases, while on the other hand – apart from isolated cases – unlike English *self-*, it is incompatible with STATES.

Other interesting cases that can be modeled with the devices of Construction Morphology concern more typically diachronic phenomena, such as the case of Modern Greek and Hungarian.

In these cases, modern $_{CC}$markers derive their constraints from motivation links that have a diachronic nature, in the sense that can be explained as depending on formal correspondences with syntactically free elements, displayed in the past functions that are no longer available to their modern counterparts.

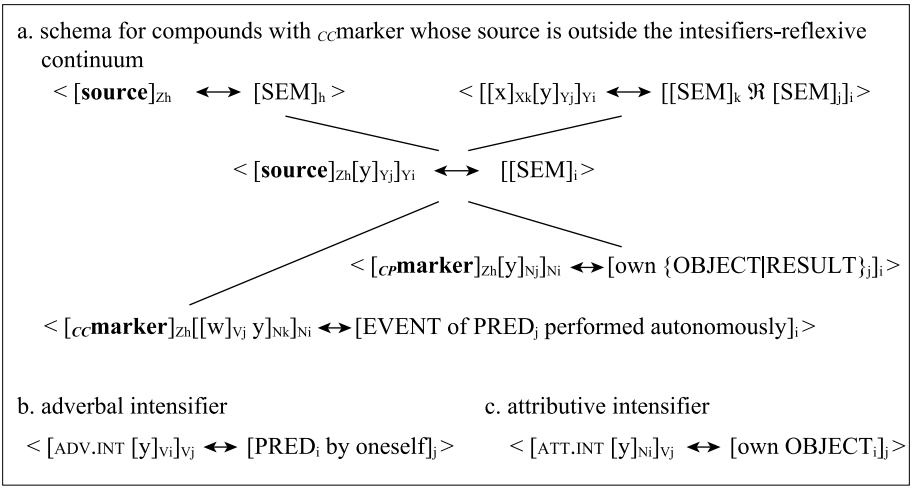


Figure 30. Compound $_{CC}$markers III × Compound $_{CC}$markers/$_{CP}$markers V

We reproduce here in Figure 16 the combination of the synchronic schema of the cases of Hungarian and Greek that we proposed at the end of Chapter 4, in section 4.4, with the schema for compounds that can have the functions of both contrastive coreference ($_{CC}$formations) and contrastive possession ($_{CP}$formations).

We can further enrich the schema with paradigmatic relations established with an antecedent historical phase, which provides the motivation that has gone lost in present times and explains the deviation of the two schemas from the tendencies concerning the sources of $_{CC}$markers observed in Chapter 4, section 4.3. In this way, the case of Modern Greek and Hungarian can be reduced to that of many other languages of the sample, as can be seen below in Figure 31, which constitutes a modified version of Figure 3 in Chapter 4. For simplicity's sake, we will represent the schema for Modern Greek *afto-* by focusing on contrastive coreference alone, and we will compare it with the syntactically free use of the adnominal intensifier of Ancient Greek *autos* (see above 6.2.1).

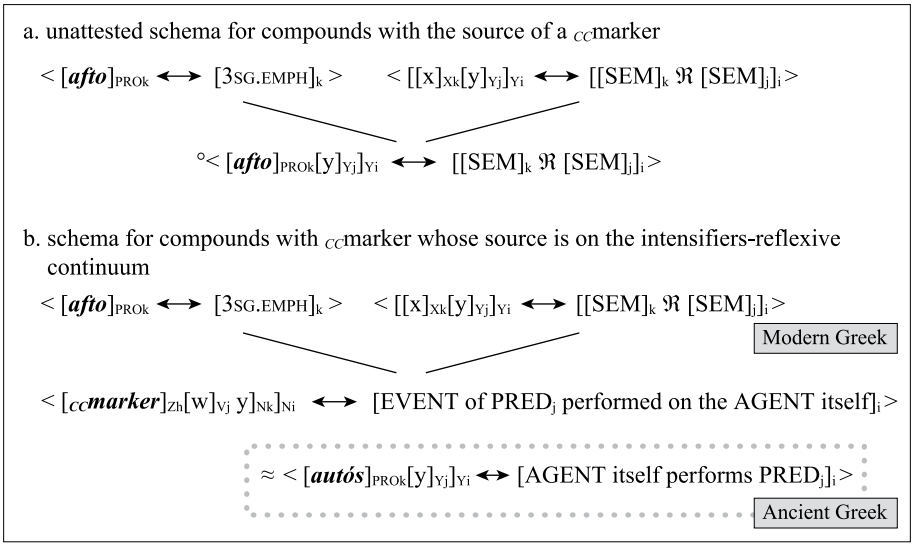


Figure 31. Compound $_{CC}$markers IIa (diachronic motivations for Modern Greek *afto-*)

It is worth mentioning that in section 4.4 above we have ascribed the case of Turkish (see sections 4.2.8 and 6.1.1) to the type "Compound $_{CC}$markers I" along with the $_{CC}$markers of all Germanic languages and of Albanian, Georgian (*tavis-* and *tav-*), and Basque. However, Turkish *öz-* was actually an exception in that class, given that *öz* does not have any function on the intensifiers-reflexive contin-

uum. The exception of Turkish $_{CC}$marker *öz-* can be explained and reduced to the aforementioned type of $_{CC}$markers if observed from the perspective of diachronic or contact-related motivations in the same vein as the cases of Modern Greek *afto-* and Hungarian *ön-*. Further research is needed here, but for illustration's needs we reproduce in Figure 32 below the schema for "Compound $_{CC}$markers I" modified to represent the case of Turkish *öz-* as supposedly motivated by the syntactic use of the adnominal intensifier of Uzbek, *o'z*.

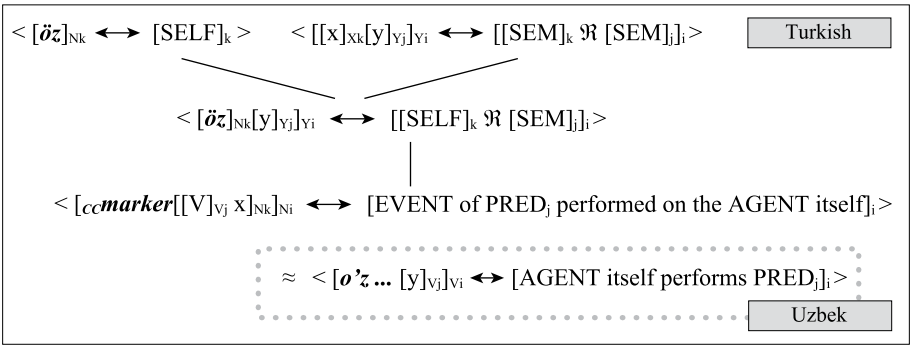


Figure 32. Compound $_{CC}$markers I (diachronic motivations for Turkish *öz-*)

In other cases, however, the change in the range of functions of the syntactically free sources of $_{CC}$markers led to a renewal of the markers themselves. This is the case of Slavic and Germanic $_{CC}$markers, whose compatibility with domains different from that of ACTIONS has progressively reduced over time at the advantage of new $_{CP}$markers. The case of Germanic and Slavic markers can be taken as evidence of a recurrent diachronic pattern by which attributive intensifiers, overlapping in a previous stage with adnominal or adverbial intensifiers, renew their form in syntax with the consequent formal renewal in word-formation as well. In this view, the schemas in the Figures presented in Chapters 4 and 5 can be ordered diachronically, and be read as different stages in the constructionalization of $_{CC}$markers. Figure 32 here below synthesizes this cline.

The cline sketched in Figure 32 is on the one hand based on the developments recognized in this chapter. On the other hand, some links are actually hypothesized, for example, the development of a strictly derivational $_{CC}$marker from compound $_{CC}$markers of type III. The only examples in the sample of $_{CC}$markers that are recognizable as derivational are in fact either due to borrowing (Romance *AUTO-*), or have never passed through a stage of compound $_{CC}$marker (Breton *em-*).

The relative position of types I and IIa/b depends on aspects concerning more whether an element of a major lexical class formally overlapping with intensifiers is available in synchrony or not . Given that the source of intensifiers is also on a grammaticalization cline as a stage preceding the development of intensifiers themselves, one might be tempted to put $_{CC}$markers of type I to the left of $_{CC}$markers of type IIa/b. However, the presence in synchrony of an element of a major lexical class formally overlapping with the intensifier in the languages of the sample is sometimes a matter of degrammaticalization (consider the emergence of *(das) Selbst* as a noun from the intensifier and not vice versa), rather than the persistence of a relation between the source of a grammatical marker and the marker itself.

The horizontal dotted curly bracket linking the upper part of the graph with its lower part, indicates that each type of compound $_{CC}$markers can potentially also constitute a $_{CP}$marker if the adnominal or adverbial intensifier is identical to the attributive intensifier (types I and IIa/b), or if an older stage of the language represented that state of affairs, but it no longer does in synchrony (type III). In this sense, we can consider that Ancient Greek, old stages of Germanic and Slavic languages, old stages of Hungarian constituted this case of identity of $_{CC}$markers and $_{CP}$markers, while belonging to type II.

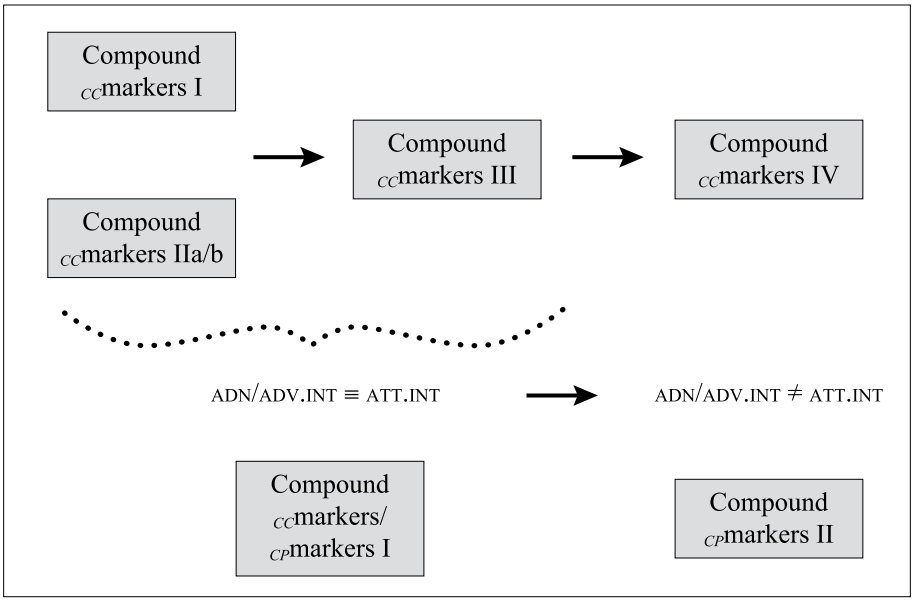


Figure 33. Diachronic relations between compound $_{CC}$markers and $_{CP}$markers.

When the emergence of new attributive intensifiers occurs, then different outcomes are possible. On the one hand, the $_{CC}$markers of Germanic and Slavic languages lost their function as $_{CP}$markers, and new attributive intensifiers motivated formations that were marked with the persisting $_{CC}$marker in earlier stages. On the other hand, we have the case of Modern Greek and Hungarian, where the $_{CC}$marker remained crystalized over the centuries, maintaining its form and its compatibility with bases on the conceptual space of POSs, while its motivation links became weaker and weaker. Consequently, they switched to type III $_{CC}$markers, but despite the emergence of new attributive intensifiers, they also remained the (main) $_{CP}$markers.

As a general conclusion to this chapter, the survey of diachronic phenomena linked to the $_{CC}$markers of European languages is a testing ground for new applications of construction morphology beyond the limits of language-specific morphological operations. The survey has shown on the one hand how diachronic developments and contact can have an important role in shaping the behaviour of morphological operations, and even in triggering their development in a language. On the other hand, the survey has stimulated the application of Construction Morphology to new unexplored territories, such as the ones of contact and of diachrony. The concept of motivation links can be fruitfully applied to these new domains, and allow for wider explanations of word-formation phenomena and for the strengthening and confirmation of generalizations and tendencies observed in synchrony.

7

CONCLUSION

At the end of this survey, which started from the contrast between Italian and German, and then widened both horizontally to encompass other European languages, and vertically to cast a glance on the past stages of some of these languages, we can draw some conclusions.

The concept of contrastive coreference, developed at the end of Chapter 1, constituted a *tertium comparationis* that was usefully exploited in Chapters 4 and 5 to provide a synchronic overview of the markers used in the languages of Europe to express it. This overview has been enriched in Chapter 6 with a series of diachronic considerations, intended to highlight the evolutionary dynamics of the current markers.

From a methodological point of view, this research, as an attempt to examine a portion of the domain of word-formation in a cross-linguistic perspective, has particularly benefited from a functionalist approach based on the identification of phenomena of interest on a fundamentally semantic basis. At least in the first phase, this approach was operationally understood in an onomasiological sense. On the other hand, also in this functionalistic-onomasiological perspective, the formal characteristics of the phenomena considered have played a primary role, not only because they are at the basis of the preliminary delimitation of the scope of the study to word-formation devices, but also because they have been central in the subsequent analysis of the operations identified in the various languages of the sample.

Also from a methodological point of view, the choice to rely almost exclusively on lexicographic sources – frequently bilingual dictionaries – has certainly limited the descriptive adequacy of the results of this study. For example, it was impossible to discern within the data of individual languages between institutionalized formations and calqued formations strictly dependent on foreign models. On a more general level, it was only possible to guess the presence of open sets of productive formations on the one hand, and on the other hand, of unproductive patterns whose products are closed sets of formations (such as the set of formations with $_{CP}$markers in Polish). However, the use of these sources has allowed us to define a picture, certainly perfectible, but rather detailed, of the variety of operations that characterize contrastive coreference in the languages of Europe.

From the point of view of the formal characteristics of the procedures, we saw in Chapter 4 how most $_{CC}$markers can be recognized as compounding elements, although frequently characterized by the use of closed-class elements such as pronouns. The *Wortbildungsbedeutung* of the operations, however, is, as we have

just mentioned, rather poor in lexical content with which the prototypical composition is typically endowed. What is remarkable about these operations is not so much the fact (observed in several points in Chapter 2) that they are placed on the borderline between compounding and derivation (as is the case, for example, for the case of affixoids), but rather that European languages mark contrastive coreference in a very regular way with operations firmly placed in a central position on the continuum. It could be objected that even affixoids are placed in the middle of the continuum and perform similar functions in different languages. However, affixoids also often have paradigmatic alternatives in the same language, which constitute more prototypical examples of affixes, while $_{CC}$markers lack prototypically affixal alternatives both within a specific language and cross-linguistically. The only exception found, Breton *em-*, appears to be a lowly productive and diagrammatically problematic case.

From the diachronic data that we were able to gather in Chapter 6, we have observed that this non-prototypicality appears to be a stable feature of $_{CC}$markers. On the other hand, what does not appear to be diachronically stable are the restrictions on compatibility with the bases of the operations. These restrictions appear to change following the emergence of new (sources of) markers that progressively erode an existing marker's productivity domain. The cases observed are mainly linked to the emergence of markers having as sources attributive intensifiers or possessive pronouns specialized in expressing contrastive possession.

The synchronic and diachronic survey of $_{CC}$markers in the languages of Europe was aimed first and foremost at describing the variety of markers attested in the languages of the sample from the point of view of their sources and of their compatibility with different classes of base words. However, parallel to this aim, a theoretical contribution was also pursued, i.e. the application of the concepts and formalism of Construction Morphology to the cross-linguistic study of a word-formation pattern. This theoretical backing allowed us to represent the $_{CC}$markers of Italian and German (*auto-* and *selbst-*, respectively) at the end of Chapter 2 by using constructional schemas with inheritance links and paradigmatic relations that show the differences in the motivation network of the word-formation patterns in the two languages. Later at the end of Chapter 4, the same theoretical devices have been exploited to represent the tendencies observed in the $_{CC}$markers of the whole language sample of European languages. The cases of Italian and German constituted two among at least four different types of $_{CC}$forma-

tions, some belonging to (non-prototypical) compounding and some to derivation. This typology expanded in Chapter 5 by the addition of two further types when discussing $_{CP}$markers and their interaction with $_{CC}$markers. The formalism of Construction Morphology has been used to formally define synchronically observable word-formation types, thus paving the way for further exploration of contrastive coreference cross-linguistically and for using Construction Morphology in the domain of morphological typology.

However, after discussing the diachrony and contact phenomena concerning European $_{CC}$markers in the first sections of Chapter 6, we engaged in an additional expansion of the scope of application of Construction Morphology. We attempted to apply its concepts and formalism beyond synchrony and across different language systems, in order to capture and represent the deviations from the tendencies concerning the sources of $_{CC}$markers observed in Chapter 4. The result of this attempt has been two-fold. On the one hand, the case of English and Italian influence on Maltese $_{CC}$formations or of the combination of English and French influence on Italian can in this way be interpreted as a kind of schema unification that operates across language systems, and that is an instance of multi-lingual language competence. On the other hand, the reinterpretation of diachronic influences of older stages of a language to more recent ones in the light of Construction Morphology has allowed us to conceive of these influences as motivation links. Following this observation, type III compound $_{CC}$markers can be seen as type I or II compound $_{CC}$markers, in which the motivation link, constituted by a paradigmatic relation with the functions of intensifier or reflexive, is not synchronic, but diachronic. This insight obtained through the application of constructional schemas and other devices of Construction Morphology strengthens the hypothesis, presented in the conclusion of Chapter 6, that the types identified in Chapter 4 and 5 can be seen as different stages on a constructionalization cline for $_{CC}$markers.

In Chapter 2, we revised a number of boundaries within word-formation and between word-formation and related fields and subfields, such as syntax and inflection. After the survey of $_{CC}$markers in the languages of Europe – a pattern that, as we have mentioned at the beginning of these conclusions, thrives at the boundary between compounding and derivation – we can stress even more the need for a more flexible approach to the definition of the subfields of morphology. What we can learn from $_{CC}$markers is that word-formation patterns displaying rather different formal features can compete and interact for expressing pretty similar

functions. In this sense, constructional schemas are a crucial device for capturing this kind of competition, while motivation links, especially paradigmatic relations with syntactic patterns, allow us to deal with phenomena that border on syntax as well – consider the case of German past participles as bases of $_{CC}$formations, or the different case forms of Estonian $_{CC}$marker *ise-*.

Regarding the perspectives of further research in this area, it is evident that the diachronic dynamics concerning $_{CC}$markers should be investigated in greater depth. For example, the kind of emphasis/intensification of the meaning of the bases that is found in some Ancient Greek formations (*autó·thermos* 'hot in itself'), but also in Old High German ones (*selb·skōnī* 'natural beauty'), should be analyzed more carefully, also considering that this kind of meaning of $_{CC}$markers appears as very rare in modern languages.

From a synchronic point of view, however, the generalizations and tendencies resulting from this study should be tested on languages outside the European area, genealogically distinct from the languages considered here, and especially typologically different from the languages of our sample. However, it would be important, in the case of the widening of the sample, especially in terms of an ampler typological variety, to refine and operationalize more precisely the criteria for distinguishing word-formation products from the products of syntax.

Another direction for further research might be the identification and analysis of the closest paradigmatic alternatives to which $_{CC}$markers can be opposed. Here we do not have in mind the alternatives with the same or similar function – the allotropes (to borrow a term of inflectional morphology) – like German *eigen-*, but the operations that constitute an alternative to $_{CC}$markers in a structured and well-defined system of traits – a paradigm in narrow sense – such as the relationship between centre and periphery that we considered as a fundamental component in the definition of the concept of contrastive coreference at the end of Chapter 1. Examples of these alternatives can be, considering, for instance, Italian word-formation, the prefix *co-* (expressing union or reciprocity) and the neoclassical combining form *etero-* (which basically reverses the pattern of expectations concerning an action by remarking an unexpected distinction of agent and patient of an action). As in this study, in the case of an analysis of these paradigmatic alternatives, an approach based on the semantic identification of the phenomena of interest also appears to be of fundamental importance in order to embrace, in a single survey, formations belonging to different classes of word-formation patterns.

8

REFERENCES

ABONDOLO, D. & VALIJÄRVI, R-L. (eds.) (2023). *The Uralic Languages. Second Edition*. Oxon, New York: Routledge.

Académia das Ciências de Lisboa (2001). *Dicionário da língua portuguesa contemporânea / da Académia das Ciências de Lisboa*. Lisboa: Verbo, ptM1.

AIKHENVALD, A. Y. (2007). Typological distinctions in word formation. In: Shopen, T. (ed.), *Language Typology and Syntactic Description, Vol. III: Grammatical Categories and the Lexicon. 2nd Edition*, 1-65. Cambridge: Cambridge University Press.

AKOBIROV, S. F., MA'RUFOVA, Z. M. & XO'JAXONOV, A. T. (1959). *O'zbekcha-ruscha lug'at*. Moscow: Xorijiy va milliy lug'atlar davlat nashriyoti, uzB1.

ALEXIADOU, A. (2014). Roots in transitivity alternations: afto/auto reflexives. In: Alexiadou, A., Borer, H. & Schäfer, F. (eds.) *The syntax of roots and the roots of syntax*, 57-80. Oxford: Oxford University Press.

AMBRAZAS, V. (ed.) (1997). *Lithuanian Grammar*. Vilnius: Baltos Lankos.

ANDERSON, S. (1992). *A-morphous Morphology*. Cambridge: Cambridge University Press.

ANGSTER, M. (2012). *Le marche di coreferenza contrastiva nella formazione delle parole. Una panoramica delle lingue d'Europa*. Unpublished PhD thesis. Università di Pavia.

ANGSTER, M. (2017). Tra formazione delle parole e sintassi. Esplorare il lessico della coreferenza contrastiva con MIDIA. In: D'Achille, P. & Grossmann, M. (eds.) *Per la storia della formazione delle parole in italiano*, 77-97. Firenze: Cesati.

AQUILINA, J. (1987-1990). *Maltese–English dictionary (2 vv.)*. Valletta, Malta: Midsea Books, mtB1.

AQUILINA, J. (1999-2000). *English–Maltese dictionary (4 vv.)*. Valletta, Malta: Midsea Books, mtB2.

ARKADIEV, P. & KLAMER, M. (2019). Morphological theory and typology. In: Audring & Masini (2019), 435-454.

ARNAUD, P. J. L. (2004a). Problématique du nom composé. In: Arnaud (2004a), 329–353.

ARNAUD, P. J. L. (ed.) (2004b). *Le nom composé. Données sur seize langues*. Lyon: Presse Universitaires.

ARONOFF, M. (1976). *Word Formation in Generative Grammar*. Cambridge: MIT Press.

ARONOFF, M. (1994). *Morphology by itself*. Cambridge: MIT Press.

ARQUIOLA, E. (2003). Morphology, argument structure and lexical semantics: The case of Spanish *auto-* and *co*-prefixation to verbal bases. *Linguistics* 41(3): 495–513.

ARTIAGOITIA, X., HUALDE, J. I. and DE URBINA, J. O. (2016). 182. Basque. In: Müller *et al.* (2016b), 3327-3348.

ASKEDAL, J. O. (2016). 141. Norwegian. In: Müller *et al.* (2016a), 2525-2554.

AUDRING, J. & MASINI, F. (eds.) (2019). *The Oxford Handbook of Morphological Theory*. Oxford: Oxford University Press.

AVRAMOVA, C. and BALTOVA, J. (2016). 167. Bulgarian. In: Müller *et al.* (2016a), 3049-3069.

BABINIÓTIS, G. D. (ed.) (2002). *Lexikó tis Néas Ellinikís glóssasghia te sosti krisi ton léxeon*. Athína: Kentro Lexikologias, elM1.

BARZ, I. (2016). 134. German. In: Müller *et al.* (2016a), 2387-2410.

BATINIĆ ANGSTER, M. (2020). *Rečenična anafora: Morfologija i sintaksa povratnih imenskih izraza u hrvatskom i talijanskom jeziku*. Zagreb: Institut za hrvatski jezik i jezikoslovlje.

BAUER, L. (1983). *English Word-Formation*. Cambridge: Cambridge University Press.

BAUER, L. (2001). Composition. In: Haspelmath *et al.* (2001).

BAUER, L. (2004). The function of word-formation and the inflection-derivation distinction. In: Aersten, H., Hannay, M. & Lyall, R. (eds.), *Words in their Places: A Festschrift for J. Lachlan Mackenzie*, 283-292. Amsterdam: Vrije Universiteit.

BAUER, L. (2005). The borderline between derivation and compounding. In: Dressler *et al.* (2005), 97–108.

BAUER, L. (2014). Concatenative derivation. In: Lieber & Štekauer (2014b), 118-135.

BAYER, M. (ed.) (2008). *Wielki słownik PONS polsko-niemiecki*. Poznań: Wydawnictwo LektorKlett / Stuttgart: Klett Sprachen, plB4.

BEEKES, R. (2010). *Etymological Dictionary of Greek.* With the assistance of *Lucien van Beek*. Leiden/Boston: Brill.

BENIGNI, V. & MASINI, F. (2009). *Compounds in Russian*. Lingue e Linguaggio, 8(2), 171–193.

BENKŐ, L. (1972). The lexical stock of Hungarian. In: *The Hungarian language*, 171–226. The Hague, Paris: Mouton.

BENKŐ, L. (ed.) (1976). *A magyar nyelv történeti-etimológiai szótára. Ö-Zs*. Buda-

pest: Akadémiai Kiadó, BENKŐ.

BERGLUND, B.-M. (ed.) (1989). *Stora svensk-tyska ordboken*. Stockholm: Esselte Studium, seB2, 4th edition.

BERGSTRØM-NIELSEN, H., LANGE, H., & LARSEN, H. V. (eds.) (1991). Dansk-tysk ordbog. København: Munksgaard, dkB1.

BEYRER, A., BOCHMANN, K., & BRONSERT, S. (eds.) (1987). *Grammatik der rumänischen Sprache der Gegenwart*. Leipzig: Verlag Enzyklopadie.

BISETTO, A. (2004). Composizione con elementi italiani. In: Grossmann, M & Rainer, F. (2004), 31-51.

BISETTO, A. (2009). Presentation: Compounds crosslinguistically. *Lingue e Linguaggio* 2/2009, 167-170.

BISETTO, A. & Scalise, S. (2005). The classification of compounds. *Lingue e Linguaggio*, 4(2), 319–332.

BLOM, C. (2005). The demarcation between morphology and syntax: a diachronic perspective on particle verbs. In: Dressler *et al.* (2005), 53–66.

BOOIJ, G. (1994). Against split morphology. In: *Yearbook of Morphology 1993*, 27–50. Dordrecht: Kluwer.

BOOIJ, G. (1996). Inherent versus contextual inflection and the split morphology hypothesis. In: *Yearbook of Morphology 1995*, 1–16. Dordrecht: Kluwer.

BOOIJ, G. (2010). *Construction Morphology*. Oxford: Oxford University Press.

BOOIJ, G. (2016). 136. Dutch. In: Müller *et al.* (2016a), 2427-2451.

BOOIJ, G. & Audring, J. (2017). Partial Motivation, Multiple Motivation: The Role of Output Schemas in Morphology. In: Booij, G. (ed). *The Construction of Words. Advances in Construction Morphology*, 59-80. Cham: Springer.

BOOIJ, G., Lehmann, C., & Mugdan, J. (eds.) (2001). *Morphologie / Morphology. Ein internationales Handbuch zur Flexion und Wortbildung*. Berlin: de Gruyter.

BORK, E. & Kaper, E. (eds.) (1986). *Tysk-dansk ordbog*. Copenhagen: Gyldendal, dkB2.

BORK, E. & Kaper, E. (eds.) (1990). *Dansk-tysk ordbog*. Copenhagen: Gyldendal, dkB3.

BOSQUE, I. & DEMONTE, V. (eds.) (1999). *Gramática Descriptiva de la Lengua Española*. Madrid: RAE-Espasa Calpe.

BOSWORTH, J. & TOLLER, N. T. (1898). *An Anglo-Saxon dictionary*. Oxford: Clarendon, BOSWORTH-TOLLER.

BOZDĚCHOVÁ, I. (2016). 158. Czech. In: Müller *et al.* (2016a), 2872-2891.

BRINCAT, J. and MIFSUD, M. (2016). 183. Maltese. In: Müller *et al.* (2016b), 3349-3366.

BUCHHOLZ, O. & FIEDLER, W. (1987). *Albanische Grammatik*. Leipzig: Enzyklopadie.

CALONGHI, F. (1950). *Dizionario latino italiano*. Torino: Rosenberg & Sellier, CALONGHI.

CASTELLA, M. (2010). *The semantics and syntax of Auto*. MA Thesis. Utrecht: Utrecht University.

CENNAMO, M. (1993). *The reanalysis of reflexives: a diachronic perspective*. Napoli: Liguori.

CIEŚLA, H., JAMROZIK, E., & KŁOS, R. (2001). *Wielki słownik włosko-polski*. Warszawa: Wiedza Powszechna, plB3.

CORNILLET, G. (2006). *Wörterbuch Bretonisch - Deutsch, Deutsch - Bretonisch*. Hamburg: Buske, brB1.

COROMINAS, J. (ed.) (1980-1991). *Diccionario crítico etimológico castellano e hispánico*. Madrid: Gredos, COROMINAS.

CORTELAZZO, M., ZOLLI, P. (1979-1988). *DELI: Dizionario etimologico della lingua italiana*. Bologna: Zanichelli, DELI.

COX, H. L. (ed.) (1983). *Groot woordenboek Duits-Nederlands*. Utrecht/Antwerpen: Van Dale Lexicographie, nlB1.

CROFT, W. (1991). *Syntactic categories and grammatical relations: the cognitive organization of information*. Chicago, London: The University of Chicago Press.

CROFT, W. (2001). *Radical Construction Grammar*. Oxford: Oxford University Press.

CROFT, W. (2003). *Typology and Universals, Second Edition*. Cambridge: Cambridge University Press.

ĆORIĆ, B. (2016). 165. Serbian. In: Müller *et al.* (2016a), 3017-3037.

CSEPREGI, M. (2023). 16. Khanty. In: Abondolo & Valijärvi (eds.), 703-752.

DAHL, O. & KOPTJEVSKAJA-TAMM, M. (eds.) (2001). *Circum-Baltic Languages. Volume I: Past and Present*. Amsterdam, Philadelphia: Benjamins.

DEANOVIĆ, M. & JERNEJ, J. (1994). *Hrvatsko-talijanski rječnik*. Zagreb: Školska knjiga, hrB2.

DEANOVIĆ, M. & JERNEJ, J. (1998). *Talijansko-hrvatski rječnik*. Zagreb: Školska knjiga, hrB3.

DELL'AGATA, G. (1972). La questione della lingua pressi i cechi: le apologie del

ceco nell'ultimo quarto del XVIII secolo. In: *Studi sulla questione della lingua presso gli slavi*. Roma: Edizioni Dell'Ateneo.

DE MAURO, T. (ed.) (1999). *Grande dizionario della lingua italiana*. Torino: UTET, GRADIT.

DESBORDES, Y. (1990). *Petite Grammaire du Breton Moderne. 2e édition augmentée*. Mouladurioù Hor Yezh.

DHRIMO, ALI & BEZHANI, H. (1996). *Deutsch-albanisches Wörterbuch [Bänden: 1. A-M, 2. L-Z]*. Wiesbaden: Harrassowitz, sqB1.

DINNEEN, P. S. (1975). *Foclóir gaedhilge agus béarla. New edition, revised and greatly enlarged, reprinted*. Dublin: Irish Texts Society, irB1.

DI SCIULLO, A. M. & WILLIAMS, E. (1987). *On the definition of word*. Cambridge, MA: MIT Press.

DI SCIULLO, M. A. (1996). Word-internal pronouns and reflexives. In: de Mulder, W.M & Tasmowski, L. (eds.) *Coherence and Anaphora*, 103–121. Amsterdam and Philadelphia: Benjamins.

DIXON, R. M. W. & AIKHENVALD, A. Y. (2001). Word: a typological framework. In: Dixon, R. M. W. & Aikhenvald, A. Y. (eds.) (2001). *Word. A cross-linguistic typology*, 1–41. Cambridge: Cambridge University Press.

DOLCETTI CORAZZA, V. (1997). *La Bibbia gotica e i bahuvrīhi*. Alessandria: Dell'Orso.

DONALIES, E. (2002). *Die Wortbildung des Deutschen. Ein Überblick*. Tübingen: Narr.

DRESSLER, W., KASTOVSKY, D., PFEIFFER, O., & RAINER, F. (eds.) (2005). *Morphology and its demarcations*. Amsterdam, Philadelphia: Benjamins.

DROSDOWSKI, G. (ed.) (1999). *Duden: das große Wörterbuch der deutschen Sprache: in zehn Bänden*. Mannheim: Dudenverlag, DUDEN.

DRYER, M. S. (2013a). Prefixing vs. Suffixing in Inflectional Morphology. In: Dryer & Haspelmath (eds.) (2013). (Available online at http://wals.info/chapter/26, Accessed on 2022-02-20.)

DRYER, M. S. (2013b). Order of Subject, Object and Verb. In: Dryer & Haspelmath (eds.) (2013). (Available online at http://wals.info/chapter/81, Accessed on 2022-02-20.)

DRYER, M. S. (2013c). Order of Adposition and Noun Phrase. In: Dryer & Haspelmath (eds.) (2013). (Available online at http://wals.info/chapter/85, Accessed on 2022-02-20.)

DRYER, M. S. (2013d). Order of Genitive and Noun. In: Dryer & Haspelmath (eds.) (2013). (Available online at http://wals.info/chapter/86, Accessed on 2022-02-20.)

DRYER, M. S. (2013e). Order of Adjective and Noun. In: Dryer & Haspelmath (eds.) (2013). (Available online at http://wals.info/chapter/87, Accessed on 2022-02-20.)

DRYER, M. S. & HASPELMATH, M. (eds.) (2013). *The World Atlas of Language Structures Online. Leipzig: Max Planck Institute for Evolutionary Anthropology*. (Available online at http://wals.info, Accessed on 2022-02-20.)

DUGAS, A. (1992). Le préfixe auto-. *Langue Française* 96: 20–29.

EDMONDSON, J. A. & PLANK, F. (1978). Great expectations: An intensive self analysis. *Linguistics and Philosophy*, 2(3), 373–413.

EISENBERG, P. (2013). *Grundriss der deutschen Grammatik Band 1: Das Wort. Unter Mitarbeit von Nanna Fuhrhop. 4., aktualisierte und überarbeitete Auflage*. Stuttgart, Weimar: Metzler.

ELLERTSSON, B. (ed.) (1993). *Íslensk-þýsk orðabók*. Reykjavík: Iðunn, isB1.

ELSEN, H. (2011). *Grundzüge der Morphologie des Deutschen*. Berlin, Boston: de Gruyter.

Esselte Studium (1993). *Tysk-svensk ordbok*. [Solna]: Esselte Studium, seB1, 2nd edition.

FAVEREAU, F. (1997). *Grammaire du breton contemporain. Yezhadur ar brezhoneg a-vremañ*. Morlaix: Skol Vreizh.

FEJES, L. (2016). 178. Permic. In: Müller *et al.* (2016b), 3260-3274.

FLEISCHER, W. (1969). *Wortbildung der deutschen Gegenwartsprache*. Leipzig: VEB Bibliographisches Institut.

FLEISCHER, W. & BARZ, I. (1995). *Wortbildung der deutschen Gegenwartsprache. 2. durchgesehene und ergänzte Auflage*. Tübingen: Niemeyer.

FLEISCHER, W. & BARZ, I. (2012). *Wortbildung der deutschen Gegenwartsprache. 4. Völlig neu bearbeitete Auflage*. Berlin, Boston: De Gruyter.

FLEURIOT, L. (1964). *Le vieux bréton. Éléments d'une grammaire*. Paris: C. Klincksieck.

FLORICIC, F. (2016). 147. French. In: Müller *et al.* (2016a), 2661-2682.

FORCELLINI, E., FURLANETTO, G., CORRADINI, F. & PERIN, J. (1864-1887). *Lexicon totius latinitatis*. Patavii: Typis Seminarii, LTL.

FRAJZYNGIER, Z. & CURL, T. S. (eds.) (1999a). *Reflexives: Forms and functions*. Amsterdam, Philadelphia: Benjamins.

FRAJZYNGIER, Z. & CURL, T. S. (eds.) (1999b). *Reciprocals: Forms and func-*

tions. Amsterdam, Philadelphia: Benjamins.
FROMM, H. (1982). *Finnische Grammatik*. Heidelberg: Winter.
GAETA, L. (2010). Synthetic compounds. With special reference to German. In: Scalise & Vogel (eds.) (2010), 219-236.
GAETA, L. (2015). 48. Restrictions in word-formation. In: Müller *et al.* (2015b), 859-875.
GAETA, L. & ANGSTER, M. (2019). Stripping paradigmatic relations out of the syntax. *Morphology* 23, 249-270.
GAETA, L. & RICCA, D. (2009). Composita solvantur: Compounds as lexical units or morphological objects? *Italian Journal of Linguistics*, 21(1), 35–70.
GAST, V., HOLE, D., KÖNIG, E., SIEMUND, P., & TÖPPER, S. (2007). *Typological Database of Intensifiers and Reflexives*. Version 2.0. http://www.tdir.org [the link redirects to https://linktype.iaa.uni-jena.de/tdir/main2.htm, accessed last time on 6th February 2022].
GENESIN, M. & MATZINGER, J. (2016). 171. Albanian. In: Müller *et al.* (2016b), 3124-3137.
GEREÑO, X. (1992). *Diccionario general castellano-euskara, euskara-castellano*. Bilbao: Amado, euB1.
GOLDBERG, A. (1995). *Constructions. A Construction Grammar Approach to Argument Structure*. Chicago: Chicago University Press.
GOLDBERG, A. (2006). *Constructions at Work. The Nature of Generalization in Language*. Oxford: Oxford University Press.
GOLDBERG, A. (2009). The nature in generalization in language. *Cognitive Linguistics*, 20(1), 93–127.
GONDAR OUBIÑA, M. T. (ed.) (2005). *PONS-Handwörterbuch: Spanisch-Deutsch*. Barcelona; Stuttgart [u.a.]: Klett, esB1.
GRANDI, N. (2002). *Morfologie in contatto. Le costruzioni valutative nelle lingue del Mediterraneo*. Milano: Franco Angeli.
GRČEVIĆ, M. (2016). 164. Croatian. In: Müller *et al.* (2016a), 2998-3016.
GRIMM, J. & GRIMM, W. (1854-1961). *Deutsches Wörterbuch*. Leipzig: Hirzel, GRIMM.
GRIMSHAW, J. (1982). On the lexical representation of Romance Reflexive Clitics. In: Bresnan, Joan (ed.): *The mental representation of grammatical relations*, 87-148. Massachusetts: MIT Publications.
GRIMSHAW, J. (1990). Reflexive Cliticization. In: Bresnan, Joan (ed.): *Argument*

structure, 152-158. Massachusetts: MIT Publications.

GROSSMANN, M. (2016). 151. Romanian. In: Müller *et al.* (2016a), 2731-2751.

GROSSMANN, M. & RAINER, F. (eds.) (2004). *La formazione delle parole in italiano*. Tübingen: Niemeyer.

GROSSMANN, M. & RAINER, F. (2009). Italian adjective-adjective compounds: Between morphology and syntax. *Italian Journal of Linguistics*, 21(1), 71–96.

GÖTZSCHE, H. (2016). 140. Danish. In: Müller *et al.* (2016a), 2505-2524.

GUEVARA, E. & SCALISE, S. (2008). Searching for Universals in Compounding. In: Scalise *et al.* (2008), 101–128.

HERCZEG, G. & JUHÁSZ, ZS. (2000). *Olasz-Magyar szótár – Vocabolario Ungherese-Italiano*. Akadémiai Kiadó, huB1.

HALLE, M. & MARANTZ, A. (1993). Distributed Morphology and the pieces of inflection. In: Kenneth L. Hale and Samuel J. Keyser (eds.). *The View from Building 20: Linguistic Essays in Honor of Sylvain Bromberger*, 111–76. Cambridge: MIT Press.

HAMMARSTRÖM, H., FORKEL, R., HASPELMATH, M. & BANK, S. (2021). *Glottolog 4.5*. Leipzig: Max Planck Institute for Evolutionary Anthropology. https://doi.org/10.5281/zenodo.5772642 (Available online at http://glottolog.org, Accessed on 2022-01-14.)

HASPEMATH, M. (1998). How young is Standard Average European? *Language Sciences* 20(3), 271-287.

HASPEMATH, M. (2003). The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In: Tomasello (2003), 211–242.

HASPEMATH, M. (2008). A frequentist explanation of some universals or reflexive marking. Linguistic Discovery, 6(1), 40–63.

HASPELMATH, M., KÖNIG, E., OESTERREICHER, W., & RAIBLE, W. (eds.) (2001). *Language typology and language universals: an international handbook*. Berlin: de Gruyter.

HATOUT, N. & NAMER, F. (2019). Paradigms in word formation: what are we up to? *Morphology* 29, 153–165.

HAY, J. & PLAG, I. (2004). What constrains possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. *Natural Language and Linguistic Theory* 22(3), 565-596.

HEINE, B. (1999). Polysemy involving reflexive and reciprocal markers in African languages. In: Frajzyngier and Curl (1999b), 1–30.

HENSCHEL, H. & KABESCH, F. (1996 (1978)). *Langenscheidts Taschenwörterbuch Tschechisch*. Berlin [u.a.]: Langenscheidt, czB1.

HERMAN, L. J. (1975). *Dictionary of Slavic word families*. New York, London: Columbia University Press.

HEWITT, B. G. (1995). *Georgian. A structural reference grammar*. Amsterdam, Philadelphia: Benjamins.

HIRVENSALO, L. (1963). *Saksalais-suomalainen sanakirja*. Porvoo: Söderström Osak., fiB1.

HOFFMANN, L. (ed.) (2007). Handbuch der deutschen Wortarten. Berlin, New York: De Gruyter.

HOHENHAUS, P. (2005). Lexicalization and institutionalization. In: Štekauer and Lieber (2005), 353–373.

HOLMES, P. & HINCHLIFFE, I. (1994). *Swedish. A Comprehensive Grammar*. London: Routledge.

HOLTON, D., MACKRIDGE, P. & PHILIPPAKI-WARBURTON, I. (2012). *Greek. A comprehensive grammar of the modern language. 2nd Edition*. London: Routledge.

HOPPE, G., KIRKNESS, A., LINK, E., NORTMEYER, I., RETTIG, W. & SCHMIDT, G. (eds.) (1987). *Deutsche Lehnwortbildung*. Tübingen: Narr.

HUALDE, J. I. & DE URBINA JON ORTIZ (eds.) (2003). *A grammar of Basque*. Berlin, New York: Mouton de Gruyter.

HÜNING, M. & BOOIJ, G. (2014). From compounding to derivation. The emergence of derivational affixes through "constructionalization". *Folia Linguistica* 48(2), 579-604.

HYSA, R. (1997). *English-Albanian*. New York: Hippocrene Books, sqB3.

IACOBINI, C. (2003). Due casi di interferenza dell'inglese sulla morfologia derivazionale dell'italiano. In: A.-V. Sullam Calimani (ed.), *Italiano e Inglese a confronto. Atti del convegno «Italiano e Inglese a confronto: problemi di interferenza linguistica», Venezia, 12-13 aprile 2002*, 43–56. Firenze: Franco Cesati Editore.

IACOBINI, C. (2004a). Composizione con elementi neoclassici. In: Grossmann and Rainer (2004), 69–95.

IACOBINI, C. (2004b). Prefissazione. In: Grossmann and Rainer (2004), 97–163.

IMBS, P. (ed.) (1970-1994). *Trésor de la langue française: dictionnaire de la langue du XIXe et du XXe siècle (1789 - 1960)*. Paris: Gallimard, TLF.

INDRIÐASON, Þ. G. (2016). 143. Icelandic. In: Müller *et al.* (2016a), 2578-2599.

JACKENDOFF, R. (1997). *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.

JACKENDOFF, R. (2002). *Foundations of Language*. Oxford: Oxford University Press.

JACKENDOFF, R. (2008). Construction after construction and its theoretical challenge. Language, 84, 8–28.

JACKENDOFF, R. & AUDRING, J. (2019). Relational Morphology in the Parallel Architecture. In: Audring and Masini (eds.) (2019), 390-408.

JAKUBÍČEK, M., KILGARRIFF, A., KOVÁŘ, V., RYCHLÝ, P. & SUCHOMEL, V. (2013). The TenTen Corpus Family. In: *7th International Corpus Linguistics Conference CL*, 125-127.

JELDEN, M. (2001). *Wörterbuch Deutsch-Georgisch Georgisch-Deutsch*. Hamburg: Buske, geB1.

JEŽEK, E. (2016). *The Lexicon. An Introduction*. Oxfod: Oxford University Press.

JÓNSSON, J. H. (2005). *Stóra orðabókin um íslenska málnotkun*. Reykjavík: JPV Útg., isM1.

JOSEPH, B. D. & PHILIPPAKI-WARBURTON, I. (1987). *Modern Greek*. London: Croom Helm.

KÁROLY, L. (2016). 186. Tatar. In: Müller *et al.* (2016b), 3398-3413.

KARPILOVSKA, I. (2016). 160. Ukrainian. In: Müller *et al.* (2016a), 2913-2931.

KAY, P. & FILLMORE, C. (1997). Grammatical Constructions and Linguistic Generalizations: The *What's X Doing Y?* Construction. *Language* 75(1), 1-33.

KEMMER, S. (1993). *The Middle Voice*. Amsterdam, Philadelphia: Benjamins.

KENESEI, I., Vago, R. M. & Fenyvesi, A. (1998). *Hungarian. A descriptive grammar*. London, New York: Routledge.

KERGE, K. (2016). 177. Estonian. In: Müller *et al.* (2016b), 3228-3259.

KIEFER, F. (2016). 181. Hungarian. In: Müller *et al.* (2016b), 3308-3326.

KILGARRIFF, A., RYCHLY, P., SMRZ, P. & TUGWELL, D. (2004). The sketch engine. In: *Proceedings of the 11th EURALEX International Congress*, 105-116.

KISCH, R. (ed.) (1989). *Dicţionar german-român*. Bucureşti: Ed. Academiei Republicii Socialiste România, roB1.

KLAJN, I. (1972). *Influssi inglesi nella lingua italiana*. Firenze: Olschki.

KLAJN, I. (2003). *Italjansko-srpski rečnik*. Beograd: Nolit, srB1.

KLAUSMANN, B. (ed.) (2009). *Taschenwörterbuch Französisch: französisch-deutsch, deutsch-französisch*. Berlin; München; Wien; Zürich [u.a.]: Langen-

scheidt, frB1.
KÖBLER, G. (ed.) (1993). *Wörterbuch des althochdeutschen Sprachschatzes*. Paderborn [u.a.]: Schoningh, KÖBLER.
KOCH, P. & MARZO, D. (2007). A two-dimensional approach to the study of motivation in lexical typology and its first application to French high-frequency vocabulary. *Studies in Language* 31: 259–291.
KOEFOED, G. & VAN MARLE, J. (2001). Productivity. In: Booij *et al.* (2001), 301-311.
KÖNIG, E. (1998). Towards a typology of intensifiers. In: Caron, B. (ed.), *Proceedings of the XVIth Congress of Linguists*. Amsterdam: Pergamon.
KÖNIG, E. (2001). Intensifiers and reflexive pronouns. In: Haspelmath et al. (2001), 747–760.
KÖNIG, E. (2011). Reflexive nominal compounds. *Studies in Language*, 35(1), 112–127.
KÖNIG, E. & GAST, V. (2006). Focused assertions of identity: A tipology of intensifiers. *Linguistic Typology*, 10, 223–276.
KÖNIG, E. & SIEMUND, P. (1999). Intensifiers and reflexives: a typological perspective. In: Frajzyngier and Curl (1999a), 41–74.
KÖNIG, E. & SIEMUND, P. (2000). The development of complex reflexives and intensifiers in English. *Diachronica*, XVII (1), 39–84.
KORDIĆ, L. (2021). *Njemački jezik u školstvu i javnom životu grada Osijeka kroz povijest*. Osijek: Pravni fakultet Sveučilišta Josipa Jurja Strossmayera u Osijeku.
KOTCHEVA, K. (2016). 142. Swedish. In: Müller *et al.* (2016a), 2554-2578.
KOVALËV, V. (ed.) (2000). *Dizionario russo-italiano, italiano-russo*. Bologna: Zanichelli, ruB2.
KRESS, B. (1982). *Isländische Grammatik*. Leipzig: VEB Verlag Enzyklopädie.
KRIŽINAUSKAS, J. & SMAGURAUSKAS, S. (2003). *Vokiečių-lietuvių kalbų žodynas*. Vilnius: Mokslo ir enciklopedijų leidybos institutas, ltB2.
KÜHNHOLD, I., PUTZER, O. & WELLMANN, H. (1978). *Deutsche Wortbildung. Vol. 3: Das Adjektiv*. Düsseldorf: Pädagogischer Verlag Schwann.
KURZ, J. (ed.) (1966-). *Slovník jazyka staroslověnského*. Praha: Nakladatelství Čekoslovenske Academie Věd, KURZ.
LAANEST, A. (1982). *Einführung in die ostseefinnischen Sprachen*. Hamburg: Buske.
LACHACHI, D. E. (2008). Zur Stellung der Halbaffigierung in der deutschen Wort-

bildung. In: Eichinger, L. M., Meliss, M. & Domínguez Vázquez, M. J. (eds.) *Wortbildung heute. Tendenzen und Kontraste in der deutschen Gegenwartsprache*, 213-230. Tübingen: Narr.

LAUČKA, A., PIESARSKAS, B. & STASIULEVIČIŪTĖ, E. (1992). *Anglų-Lietuvių kalbų žodynas*. Vilnius: Mokslas, ltB1.

LEFER, M.-A. & CARTONI, B. (2011). Prefixes in contrast: Towards a meaning-based contrastive methodology for lexical morphology. *Languages in Contrast*, 11(1), 87–105.

LEUSCHNER, T., MORTELMANS, T. & DE GROODT, S. (eds.) (2005). *Grammatikalisierung im Deutschen*. Berlin: de Gruyter.

LEVIN, B. (1993). *English verb classes and alternations: a preliminary investigation*. Chicago, London: The University of Chicago Press.

LEWIS, G. (1999). *The Turkish language reform. A catastrophic success*. Oxford: Oxford University Press.

LEWIS, H. & PEDERSEN, H. (1961). *A concise comparative Celtic grammar*. Göttingen: Vandenhoeck & Ruprecht.

LIEBER, R. (2014). Methodological Issues in Studying Derivation. In: Lieber, R. & Štekauer, P. (eds.) (2014), 84-94.

LIEBER, R. & ŠTEKAUER, P. (2009a). Introduction: Status and definition of Compounding. In: Lieber, R. & Štekauer, P. (eds.) (2009b), 3-18.

LIEBER, R. & ŠTEKAUER, P. (eds.) (2009b). *The Oxford Handbook of Compounding*. Oxford: Oxford University Press.

LIEBER, R. & ŠTEKAUER, P. (2014a). Universals in Derivation. In: Lieber, R. & Štekauer, P. (eds.) (2014b), 778-786.

LIEBER, R. & ŠTEKAUER, P. (eds.) (2014b). *The Oxford Handbook of Derivational Morphology*. Oxford: Oxford University Press.

LINGEA (2006). *Německo-český, česko-německý velký slovník: ...nejen pro překladatele*. Brno: Lingea, czB2.

LO CASCIO, V. (ed.) (2001). *Il dizionario neerlandese. Van Dale. 2 vv*. Utrecht, Antwerpen: Van Dale Lexicografie; Bologna: Zanichelli, nlB2.

LOHDE, M. (2006). *Wortbildung des modernen Deutschen. Ein Lehr- und Übungsbuch*. Tübingen: Narr.

LOTH, J. (1982). *Vocabulaire vieux-breton*. Geneve: Slatkine [reissue of the 1884 edition published by Champion, Paris].

LÖTZSCH, R. (ed.) (1991). *Deutsch-Russisches Wörterbuch*. Berlin: Akademie-Ver-

lag, ruB3.
LYONS, J. (1968). *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press.
MACAULAY, D. (ed.) (1992a). *The Celtic Languages*. Cambridge: Cambridge University Press.
MACAULAY, D. (1992b). The Scottish Gaelic Language. In: MacAulay (1992a), 137-248.
MACURA, P. (1999). *Elsevier's Russian-English dictionary*. Amsterdam [u.a.]: Elsevier, ruB4, 2nd edition.
MAIDEN, M., SMITH, J. C., GOLDBACH, M. & HINZELIN, M. (eds.) (2011). *Morphological Autonomy: Perspectives from Romance Inflectional Morphology*. Oxford: Oxford University Press.
MAIER, B. (2011). *Wörterbuch Schottisch-Gälisch – Deutsch, Deutsch – Schottisch-Gälisch*. Hamburg: Buske, scB1.
MANOVA, S. (2015). 54. Closing suffixes. In: Müller *et al.* (2015b), 957-972.
MARELLO, C. (1996). *Le parole dell'italiano*. Bologna: Zanichelli.
MASINI, F. & AUDRING, J. (2019). Construction Morphology. In: Audring & Masini (2019), 365-389.
MARTÍNEZ RUBIO, E. (2006). *Euskara-alemana hiztegia. Wörterbuch deutsch-baskisch*. Donostia (Gipuzkoa): Elkar, euB2.
MATHIASSEN, T. (1996). *A short grammar of Lithuanian*. Columbus, Ohio: Slavica Publications.
MEA, G. (1989-1990). *O dicionario portugues*. Porto: Porto Editora [Bologna: Zanichelli], ptB1.
MEISELS, W. (1996). *Podręczny słownik polsko-włoski*. Warszawa: Wiedza Powszechna, plB2.
MIGLIORINI, B. (1963). *Saggi sulla linguistica del Novecento*. Firenze: Sansoni.
MIRA MATEUS, M. H. (ed.) (2003). *Gramática da Língua Portuguesa*. Lisboa: Editorial Caminho.
MONACHESI, P. (1999). *A lexical approach to Italian cliticization*. Stanford: CSLI.
MONTERMINI, F. (2008). *Il lato sinistro della morfologia*. Milano: F. Angeli.
MONTERMINI, F. (2010). Units in compounding. In: Scalise and Vogel (2010), 77–92.
MONTERMINI, F. (2019). Later generative grammar and beyond: Lexicalism. In: Audring & Masini (2019), 122-142.

MORAN, A. V. (ed.) (1945). *Türkçe-İngilizce sözlük*. İstanbul: Milli Eğitim Bakanlığı, tkB1.

MORAN, A. V. & MORAN, Y. (eds.) (1985). *Büyük Türkçe-İngilizce sözlük*. İstanbul: Adam, tkB2.

MORAVCSIK, E. (1972). Some cross-linguistic generalizations about intensifier constructions. *Chicago Linguistic Society* 8, 271–277.

MÜLLER, P. O., OHNHEISER, I., OLSEN, S. & RAINER, F. (eds.) (2015a). *Word-Formation. An International Handbook of the Languages of Europe. Vol. 1*. Berlin, München, Boston: De Gruyter Mouton.

MÜLLER, P. O., OHNHEISER, I., OLSEN, S. & RAINER, F.(eds.) (2015b). *Word-Formation. An International Handbook of the Languages of Europe. Vol. 2*. Berlin, München, Boston: De Gruyter Mouton.

MÜLLER, P. O., OHNHEISER, I., OLSEN, S. & RAINER, F. (eds.) (2015c). *Word-Formation. An International Handbook of the Languages of Europe. Vol. 3*. Berlin, München, Boston: De Gruyter Mouton.

MÜLLER, P. O., OHNHEISER, I., OLSEN, S. & RAINER, F. (eds.) (2016a). *Word-Formation. An International Handbook of the Languages of Europe. Vol. 4*. Berlin, München, Boston: De Gruyter Mouton.

MÜLLER, P. O., OHNHEISER, I., OLSEN, S. & RAINER, F. (eds.) (2016b). *Word-Formation. An International Handbook of the Languages of Europe. Vol. 5*. Berlin, München, Boston: De Gruyter Mouton.

MUTHMANN, G. (1994). *Doppelformen in der deutschen Sprache der Gegenwart*. Tübingen: Narr.

MUTZ, K. (2003). Le parole complesse in "auto" nell'italiano di oggi. In: N. Maraschio and T. Poggi Salani (eds.), *Atti del XXXIV Congresso Internazionale di Studi della SLI. Firenze, 19-21 novembre 2000*, 649–66. Roma: Bulzoni.

MUTZ, K. (2004). Zur Argumentstruktur der deverbalen Ableitungen von *auto-*. In: Hummel, M. & Kailuweit, R. (eds.), *Semantische Rollen in der Romania. Tübinger Beiträge zur Linguistik 472*, 355–374. Tübingen: Narr.

MUTZ, K. (2009). Polisemia del prefisso auto- in unità lessicali complesse dell'italiano. In: Burr, E. (ed.) *Tradizione e innovazione, Atti del VI convegno SILFI*. Firenze: Franco Cesati.

NAGÓRKO, A. (2016). 156. Polish. In: Müller *et al.* (2016a), 2831-2852.

NAUGHTON, J. & VAN KUNES, K. (2021). *Czech. An essential grammar*. Abingdon, New York: Routledge.

NAVICKAITĖ-KLIŠAUSKIENĖ, A. (2016). 170. Latvian. In: Müller *et al.* (2016b), 3107-3123.

NEWMARK, L. (ed.) (1998). *Albanian-English Dictionary*. Oxford: University Press, sqB2.

Ó CURNÁIN, B. (2016). 154. Irish. In: Müller *et al.* (2016a), 2782-2810.

O'DÓNAILL, N. (1998). *Foclóir Gaeilge-Béarla*. Baile Átha Cliath: Oifig an tSoláthair, irB2.

OLSEN, S. (2014). Delineating derivation and compounding. In Lieber & Štekauer (2014b), 26-49.

OLSEN, S. (2015). 20. Composition. In: Müller *et al.* (2015a), 364-386.

OREL, V. (1998). *Albanian etymological dictionary*. Leiden, Boston, Köln: Brill.

ORQUEDA, V., ARRIAGADA, S., TORO, F. (2020). Spanish [*auto* + V + *se*] constructions. *Folia Linguistica* 54(3), 615-645.

ORTNER, HANSPETER & ORTNER, L. (1984). *Zur Theorie und Praxis des Kompositaforschung: Mit einer ausführlichen Bibliographie*. Tübingen: Narr.

Oxford University Press (2012). *OED Online*. Oxford University Press. http://www.oed.com/.

PALAZZI, F. & FOLENA, G. (1992). *Dizionario della lingua italiana*. Torino: Loescher Editore, PALAZZI-FOLENA.

PAUL, H. (1881-1998). *Mittelhochdeutsche Grammatik*. Tübingen: Niemeyer, 24th edition.

PAVLOVIĆ, N. & RADOJIČIĆ, S. (2006). *Italjansko-srpski, srpsko-italjanski rečnik*. Beograd: Jasen, srB2.

PERLMUTTER, D. M. (1988). The Split Morphology Hypothesis: Evidence from Yiddish. In: *Theoretical Morphology*, 79–100. San Diego etc.: Academic Press.

PICCHIO, R. (1980). Lo slavo ecclesiastico. In: Picchio, R. (ed.). *Letteratura della Slavia ortodossa* (IX-XVIII sec.). Bari: Dedalo.

PINTO, I. (2016). 149. Sardinian. In: Müller *et al.* (2016a), 2693-2712.

PITKÄNEN-HEIKKILÄ, K. (2016). 176. Finnish. In: Müller *et al.* (2016b), 3209-3228.

PLAG, I. (2016). 135. English. In: Müller *et al.* (2016a), 2411-2427.

PLANK, F. & FILIMONOVA, E. (2000). The Universals Archive: A Brief Introduction for Prospective Users. *Sprachtypologie und Universalienforschung* 53, 109-123.

PLANK, F., MAYER, T., MAYORAVA, T. & FILIMONOVA, E. (n.d.) *The Universals Archive*. http://typo.uni-konstanz.de/archive/ [the link redirects to https://typo.uni-konstanz.de/rara/, accessed last time on 27th November 2023].

POKORNY, J. (1959). *Indogermanisches etymologisches Wörterbuch*. München: Francke, POKORNY.

PÖLL, B. (2016). 144. Portuguese. In: Müller *et al.* (2016a), 2600-2619.

PRESS, I. (1986). *A Grammar of Modern Breton*. Berlin, New York, Amsterdam: Mouton de Gruyter.

PUDDU, N. (2005). *Riflessivi e intensificatori: greco, latino e le altre lingue indoeuropee*. Pisa: ETS.

PUDDU, N. (2010). Considerazioni di tipologia morfologica sul pronome riflessivo in greco. In: Putzu, I., Paulis, G., Nieddu, G. & Cuzzolin, P. (eds.) *La morfologia del greco tra tipologia e diacronia*, 385-405. Milano: Franco Angeli.

RAFFELSIEFEN, R. (2015). 50. Phonological restrictions on English word-formation. In: Müller *et al.* (2015b), 894-917.

RAINER, F. (1993). *Spanische Wortbildungslehre*. Tübingen: Niemeyer.

RAINER, F. (2016a). 145. Spanish. In: Müller *et al.* (2016a), 2620-2640.

RAINER, F. (2016b). 150. Italian. In: Müller *et al.* (2016a), 2712-2731.

RAMAT, P. (2010). The (early) history of linguistic typology. In: Song (2010), 9-24.

RAUN, A. & SAARESTE, A. (1965). *Introduction to Estonian linguistics*. Wiesbaden: Harrassowitz.

RAYFIELD, D. (ed.) (2006). *A comprehensive Georgian-English dictionary*. London: Garnett press, geB2.

REY, A. (ed.) (1995). *Dictionnaire historique de la langue française*. Paris: Dictionnaires Le Robert, REY.

REY-DEBOVE, J. & REY, A. (eds.) (2000). *Le Nouveau Petit Robert*. Paris: Dictionnaires Le Robert, NOUVEAU PETIT ROBERT.

RICCA, D. (2004). Aggettivi deverbali. In: Grossmann & Rainer (2004), 419–444.

RICCA, D. (2005). Cumulative exponence involving derivation: Some patterns for an uncommon phenomenon. In: Dressler *et al.* (eds.) (2005), 197-213.

ROBERT, P. & REY-DEBOVE, J. (2007). *Le nouveau Petit Robert: dictionnaire alphabétique et analogique de la langue française*. Paris: Dictionnaires Le Robert, 2007, frM1, 40th edition.

ROCCI, L. (ed.) (1998). *Vocabolario greco-italiano*. Roma: Dante Alighieri, ROCCI.

RÖMER, C. (2006). *Morphologie der deutschen Sprache*. Tübingen: A. Francke.

RUBINO, C. (2013). Reduplication. In: Dryer & Haspelmath (eds.) (2013). (Available online at http://wals.info/chapter/27, Accessed on 2022-02-20.)

RUMELHART, D. E. (1980). Schemata: the building blocks of cognition. In: Spiro,

R., Bruce, B. & Brewer, W. (eds.), *Theoretical Issues in Reading Comprehension*, 33-58. Mahway: Erlbaum Associates.
RUSSELL, Paul. (2016). 153. Welsh. In: Müller *et al.* (2016a), 2769-2781.
SAAGPAAK, P. F. (1982). *Eesti-inglise sõnaraamat. Estonian-English dictionary*. New Haven-London: Yale University Press, eeB1.
SABATINI, F. & COLETTI, V. (1997). *DISC compact. Dizionario italiano. Versione 1.1*. Firenze: Giunti Multimedia, DISC.
SALTARELLI, M. (1988). *Basque*. London: Routledge.
SÁNCHEZ, A. (ed.) (2001). *Gran diccionario de uso del español actual: basado en el Corpus lingüístico Cumbre*. Alcobendas, Madrid: Sociedad General Española de Librería, esM1.
SCALISE, S. (1983). *Morfologia lessicale*. Padova: Clesp.
SCALISE, S. (1994). *Morfologia*. Bologna: Il Mulino.
SCALISE, S. & BISETTO, A. (2009). The classification of compounds. In: Lieber and Štekauer (2009), 34–53.
SCALISE, S., MAGNI, E., & BISETTO, A. (eds.) (2008). *Universals of Language Today*. Amsterdam: Springer.
SCALISE, S. & VOGEL, I. (eds.) (2010). Cross-disciplinary issues in compounding. Amsterdam: Benjamins.
SCHLADT, M. (1999). The typology and grammaticalization of reflexives. In: Frajzyngier and Curl (1999a), 103–124.
SCHMIDT, G. D. (1987). Das Affixoid: Zur Notwendigkeit und Brauchbarkeit eines beliebten Zwischenbegriffs der Wortbildung. In: Hoppe *et al.* (1987).
SENN, A. (1966). *Handbuch der Litauischen Sprache. Band I: Grammatik*. Heidelberg: Carl Winter Universitätsverlag.
SFORZA, I. (2007). Materiali per una storia dei composti di αὐτο-. *Studi e saggi linguistici*, XLV, 99–115.
SIEMUND, P. (2007). Reflexivum. In: Hoffmann (2007), 707–726.
SIGURJÓNSDÓTTIR, S. & HYAMS, N. (1992). Reflexivization and logophoricity: evidence from the acquisition of Icelandic. *Language Acquisition*, 2(4), 359–413.
SILVET, K. J. (1989). *An English-Estonian dictionary – Inglise-eesti sõnaraamat*. Tallinn: Valgus, eeB2.
SIMPSON, J. A. & Weiner, E. (eds.) (1989). *The Oxford English Dictionnary, Second Edition*. Oxford: Clarendon, OED.
SKIBICKI, M. (2007). *Polnische Grammatik*. Hamburg: Buske.

SONG, J. (ed.) (2010). *The Oxford Handbook of Linguistic Typology*. Oxford: Oxford University Press.

STANKOVIĆ, B. (ed.) (1998). *Rusko-sprski rečnik. Drugo, ispravljeno i dopunjeno izdanje*. Novi Sad: Budućnost.

Steinarsson, H. (ed.) (2008). *PONS Kompaktwörterbuch Deutsch-Isländisch*. Stuttgart: PONS, 2008, isB2.

ŠTEKAUER, P. & LIEBER, R. (eds.) (2005). *Handbook of Word-Formation*. Dordrecht: Springer.

ŠTEKAUER, P., VARELA, S. & KÖRTVÉLYESSY, L. (2012). *Word-Formation in the Wolrd's Languages. A Typological Survey*. Cambridge: Cambridge University Press.

STEVENS, C. M. (2005). Revisiting the affixoid debate: On the grammaticalization of the word. In: Leuschner *et al.* (2005), 71–83.

STRAMLJIČ BREZNIK, I. (2016). 163. Slovene. In: Müller *et al.* (2016a), 2979-2998.

STREITBERG, W. (ed.) (2000). *Die gotische Bibel*. Heidelberg: Winter.

STUDŽIA, B. (2016). 169. Lithuanian. In: Müller *et al.* (2016b), 3090-3106.

Svenska Akademien (2009). *Svensk ordbok. Utgiven av Svenska Akademien*. Stockholm: Norstedts, seM1.

SZYMANEK, B. (2009). IE. Polish. In: Lieber and Štekauer (2009), 464–477.

ten Hacken, P. (2001). Derivation and compounding. In: Booij *et al.* (2001), 349–360.

TERNES, E. (2016). 152. Breton. In: Müller *et al.* (2016a), 2752-2768.

TESI, R. (1994). *Dal greco all'italiano. Studi sugli europeismi lessicali di origine greca dal Rinascimento a oggi*. Firenze: Le Lettere.

THOMAS, R. J., BEVAN, G. A., & DONOVAN, P. J. (eds.) (1967-2002). *Geiriadur Prifysgol Cymru: a dictionary of the Welsh language (4 volumes)*. Caerdydd: Gwasg Prifysgol Cymru, cyM1.

THOMAS, A. R. (1992). The Welsh Language. In: MacAulay (1992a), 251-345.

THORNTON, A. M. (2011). Overabundance (Multiple Forms Realizing the Same Cell) A Non-canonical Phenomenon in Italian Verb Morphology. In: Maiden *et al.* (2011), 358-381.

TOMASELLO, M. (ed.) (2003). *The new psychology of language, vol. 2*. Mahwah, NJ: Lawrence Erlbaum.

TSOPANOGLOU, A. (ed.) (1996). *Dizionario Greco Moderno-Italiano Italiano-Greco Moderno*. Bologna: Zanichelli–Perugia Edizioni, elB1.

TURCHI, P. M. (ed.) (1994). *Íslensk-ítölsk orðabók*. Reykjavík: Iðunn, isB3.
TUTSCHKE, G. (1999). *Langenscheidts Taschenwörterbuch Kroatisch*. Berlin [u.a.]: Langenscheidt, hrB1.
ULUHANOV, I. S. (2016). 162. Russian. In: Müller *et al.* (2016a), 2953-2978.
van Marle, J. (2001). Paradigmatic and syntagmatic relations. In: Booij *et al.* (2001), 225–234.
VARELA, S. & MARTIN GARCÍA, J. (1999). La prefijación. In: Bosque and Demonte (1999), 4993–5038.
VASIĽOVÁ, M. (2017). *Compounds and Identity. A Cross-Linguistic Study of SELF-compounds*. Unpublished PhD thesis, Utrecht University and Pavol Jozef Šafárik University in Košice.
VASMER, M. (1964-1973). *Etimologičeskij slovar' russkogo jazyka*. Moscow: Izdatel'stvo "Progress".
VELUPILLAI, V. (2012). *An Introduction to Linguistic Typology*. Amsterdam, Philadelphia: Benjamins.
VIDRAŞCU, A. (ed.) (2006). *Noul dicţionar universal al limbii române*. Bucuresti [u.a.]: Ed. Litera Internaţional, roM1.
VIRTANEN, S. & HORVÁTH, C. (2023). 15. Mansi. In: Abondolo & Valijärvi (eds.), 665-702.
VJATERE, L. & ANDERSONE, I. (2007). *Vācu-latviešu – latviešu-vācu vārdnīca*. Rīgā: Avots, lvB1.
WAGNER-NAGY, B. (2016). 175. Nenets. In: Müller *et al.* (2016b), 3197-3209.
WALDE, A. & HOFMANN, J. B. (eds.) (1965-1982). *Lateinisches etymologisches Wörterbuch*. Heidelberg: Winter, WALDE-HOFMANN.
WALEWSKI, S. (2002). *Langenscheidts Taschenwörterbuch Polnisch*. Berlin [u.a.] : Langenscheidt, plB1.
WALEWSKI, S. & WEDEL, E. (1994). *Langenscheidts Taschenwörterbuch Russisch*. Berlin [u.a.]: Langenscheidt, ruB1.
WENDT, H. F., KORNRUMPF, H.-J., & TURAN, T. (eds.) (2009). *Langenscheidt Taschenwörterbuch Türkisch*. Berlin, München: Langenscheidt, tkB3.
WILKENS, J. (2016). 184. Turkish. In: Müller *et al.* (2016b), 3368-3385.

Lexicographic sources: Abbreviations

INDO-EUROPEAN

POKORNY
: Pokorny (1959)

LATIN

CALONGHI
: Calonghi (1950)

LTL Forcellini *et al.* (1864-1887)

WALDE-HOFMANN
: Walde and Hofmann (1965-1982)

ITALIAN

DELI Cortellazzo and Zolli (1979-1988)

DISC Sabatini and Coletti (1997)

GRADIT De Mauro (1999)

PALAZZI-FOLENA
: Palazzi and Folena (1992)

FRENCH

frB1 Klausmann (2009)

frM1 Robert and Rey-Debove (2007)

REY Rey (1995)

NOUVEAU PETIT ROBERT
: Rey-Debove and Rey (2000)

TLF Imbs (1970-1994)

PORTUGUESE

ptB1 Mea (1989-1990)

ptM1 Académia das Ciências de Lisboa (2001)

ROMANIAN

roB1 Kisch (1989)

roM1 Vidraşcu (2006)

SPANISH

COROMINAS
: Corominas (1980-1991)

esB1 Gondar Oubiña (2005)

esM1 Sánchez (2001)

GERMAN

DUDEN
: Drosdowski (1999)

GRIMM
: Grimm and Grimm (1854-1961)

KÖBLER
: Köbler (1993)

ANGLO-SAXON

BOSWORTH-TOLLER
: Bosworth and Toller (1898)

ENGLISH

OED Simpson and Weiner (1989)

DUTCH

nlB1 Cox (1983)

nlB2 Lo Cascio (2001)

Danish
dkB1 Bergstrøm-Nielsen *et al.* (1991)
dkB2 Bork and Kaper (1990)
dkB3 Bork and Kaper (1986)

Icelandic
isB1 Ellertsson (1993)
isB2 Steinarsson (2008)
isB3 Turchi (1994)
isM1 Jónsson (2005)

Swedish
seB1 Esselte Studium (1993)
seB2 Berglund (1989)
seM1 Svenska Akademien (2009)

Old Church Slavonic
Kurz Kurz (1966-)

Croatian
hrB1 Tutschke (1999)
hrB2 Deanović and Jernej (1994)
hrB3 Deanović and Jernej (1998)

Serbian
srB1 Klajn (2003)
srB2 Pavlović and Radojičić (2006)

Czech
czB1 Henschel and Kabesch (1978)
czB2 Lingea (2006)

Polish
plB1 Walewski (2002)
plB2 Meisels (1996)
plB3 Cieśla *et al.* (2001)
plB4 Bayer (2008)

Russian
ruB1 Walewski and Wedel (1994)
ruB2 Kovalëv (2000)
ruB3 Lötzsch (1991)
ruB4 Macura (1999)

Irish Gaelic
irB1 Dinneen (1975)
irB2 O'Dónaill (1998)

Scottish Gaelic
scB1 Maier (2011)

Welsh
cyM1 Thomas *et al.* (2002)

Breton
brB1 Cornillet (2006)

Lithuanian
ltB1 Laučka *et al.* (1992)
ltB2 Križinauskas and Smagurauskas (2003)

Latvian
lvB1 Vjatere and Andersone (2007)

Ancient Greek
Rocci Rocci (1998)

Modern Greek
elB1 Tsopanoglou (1996)

elM1 Babiniótis (2002)

Albanian

sqB1 Dhrimo and Bezhani (1996)
sqB2 Newmark (1998)
sqB3 Hysa (1997)

Finnish

fiB1 Hirvensalo (1963)

Estonian

eeB1 Saagpaak (1982)
eeB2 Silvet (1989)

Hungarian

huB1 Herczeg and Juhász (2000)
Benkő
Benkő (1976)

Turkish

tkB1 Moran (1945)
tkB2 Moran and Moran (1985)
tkB3 Wendt *et al.* (2009)

Uzbek

uzB1 Akobirov *et al.* (1959)

Maltese

mtB1 Aquilina (1987-1990)
mtB2 Aquilina (1999-2000)

Georgian

geB1 Jelden (2001)
geB2 Rayfield (2006)

Basque

euB1 Gereño (1992)
euB2 Martínez Rubio (2006)

List of tables

List of figures

https://morepress.unizd.hr/books/index.php/press