



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

ARCHIVIO ISTITUZIONALE DELLA RICERCA

Alma Mater Studiorum Università di Bologna Archivio istituzionale della ricerca

The fate of 'pseudo-' words. A contrastive corpus-based analysis

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:

Availability:

This version is available at: <https://hdl.handle.net/11585/953564> since: 2024-01-20

Published:

DOI: <http://doi.org/10.1075/lic.22003.van>

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>).
When citing, please refer to the published version.

(Article begins on next page)

This is the final peer-reviewed accepted manuscript of:

Van Goethem, Kristel, Muriel Norde, and Francesca Masini. "The fate of 'pseudo-words': A contrastive corpus-based analysis." *Languages in Contrast* (2023).

The final published version is available online at:

<https://dx.doi.org/10.1075/lic.22003.van>

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (<https://cris.unibo.it/>)

When citing, please refer to the published version.

The fate of ‘pseudo-’ words: a contrastive corpus-based analysis

Kristel Van Goethem, Muriel Norde and Francesca Masini

F.R.S.-FNRS & Université catholique de Louvain (Belgium) | Humboldt
Universität zu Berlin (Germany) | Alma Mater Studiorum – Università di Bologna
(Italy)

The present study examines the fate of the neoclassical combining form *pseudo-* in eight European languages, belonging to Germanic (Danish, Dutch, English, German and Swedish) and Romance (French, Italian, Spanish). In order to gain a better understanding of the synchronic morphological behaviour and productivity of *pseudo-* words in these languages, we carry out a cross-linguistic corpus analysis and compare the morphological and distributional properties of *pseudo-*. We also analyse its debonding behaviour and categorical flexibility in the set of languages and correlate this property with its productivity. The results of the corpus study are discussed against the typological background of the so-called Germanic and Romance Sandwiches.

Keywords: neoclassical compounds, combining forms, morphological cohesion, corpus linguistics, Danish/Dutch/English/French/German/Italian/Spanish/Swedish

1 Introduction

In this study, we compare the synchronic integration of the morpheme *pseudo-* in Germanic and Romance. Its historical development makes *pseudo-* a particularly interesting topic for a cross-linguistic analysis. *Pseudo-* is a case of morphological ‘matter borrowing’ (Seifart, 2015; Gardani, 2020): originally a compounding element in Ancient Greek (e.g. *pseudologia* ‘a false speech’), *pseudo-* was borrowed into a wide variety of European languages and synchronically combines with various types of native bases.

To the best of our knowledge, large-scale contrastive studies that compare the synchronic use of a cognate neoclassical combining form have not been published to date. Interestingly, however, it could reveal the (possibly divergent) morphological features and semantic developments of these formatives, depending on the language-specific properties of each language. The aim of this study is to fill this gap by concentrating on the fate of *pseudo-* in eight European languages (Danish, Dutch, English, French, German, Italian, Spanish and Swedish), with a specific focus on its morphological properties and productivity. We will also sketch a general semantic profile of *pseudo-*, but a detailed analysis of its semantic developments is beyond the scope of this paper.

In order to compare the fate of *pseudo-* words cross-linguistically, this study provides an extensive comparative corpus-based analysis of the synchronic use of *pseudo-* while considering the morphological properties specific to each of the receiving languages.

More specifically, the research aims are as follows:

1. To compare the integration and behaviour of *pseudo-* within the Germanic and Romance languages, especially with respect to its morphological properties, frequency and productivity;
2. To investigate the relationship between the aforementioned properties on the one hand and ‘debonding’ and categorical flexibility of *pseudo-* on the other.

We embed our findings in a typological framework known as the Germanic and Romance ‘Sandwiches’ (see, e.g., van Haeringen, 1956; Hüning *et al.*, 2006; Lamiroy, 2011; König and Gast, 2018), which allows us to formulate research questions on the potentially different use and behaviour of *pseudo-* in each of the languages.

The outline of this paper is as follows. In Section 2, we introduce formations with *pseudo-* and describe its etymology, morphological status and semantic profile. Section 3 presents the typological framework underlying our contrastive study, focusing on the clines that have been observed in the Germanic and Romance languages, and the relevance of morphological factors such as cohesion

and inflection for the morphological behaviour of word formations in Germanic and Romance. Based on this framework, the research questions of our study are formulated at the beginning of Section 4. Subsequently, this section delves into the cross-linguistic corpus analysis of *pseudo-*, with a main focus on its morphological properties, frequency and productivity. Section 5 reports on the findings related to the debonding and categorical flexibility of *pseudo-* and discusses the correlation between debonding and productivity found in the eight languages under study. Finally, Section 6, presents the conclusions of our study and gives an outlook to follow-up research questions.

2 *Pseudo-*: Morphological and semantic properties

2.1 Etymology and morphological status of *pseudo-*

According to the lexicographical sources *Etymonline* and *OED* (s.v. *pseudo-*)¹, the morpheme *pseudo-* has been borrowed from Greek *pseudo-*, which is derived from the adjective *pseudēs* ‘false; falsely’ or the noun *pseudos* ‘falsehood, untruth, a lie’, which both derive from the Greek verb *pseudein* ‘to deceive, to cheat’. Bound *pseudo-* can be traced back to compound nouns and adjectives in ancient Greek (e.g. *pseudodidaskalos* ‘false teacher’, *pseudologia* ‘a false speech’, *pseudoparthenos* ‘pretended virgin’, *pseudológos* ‘speaking falsely’). Some of these Greek nouns and adjectives were adopted into classical Latin, especially in terms of natural history (e.g. *pseudosphēx* ‘false wasp’), and in post-classical Latin, *pseudo-* started to combine with Latin bases (e.g. *pseudoflavus* ‘yellowish’).

In English (*Etymonline* and *OED* (s.v. *pseudo-*)), *pseudo-* is first attested in the early 15th century, both in adaptations of post-classical Latin words or their Hellenistic Greek etymons (e.g. *pseudoprophet*) and in adaptations of Latin words from Christian contexts (e.g. *pseudo-priest*). *Pseudo-* starts to combine with

¹https://www.etymonline.com/word/pseudo-#etymonline_v_2781 and <https://www.oed-com>

English native words in later Middle English with the meaning of ‘false, hypocritical’ (e.g. *pseudochristian*) and becomes productive in the course of the 19th century in both nominal and adjectival formations, particularly in scientific terminology. By that time English formations in *pseudo-* are not restricted to words that derive from Greek or Latin.

We find parallel historical developments in the other Germanic and the Romance languages included in this study. Initially, Greek or Latin formations are borrowed as a whole (e.g. *pseudonym*). Later, *pseudo-* comes to be combined with native words, mostly pertaining to (spurious) science or politics (e.g. Dutch *pseudodeskundige* ‘pseudo-expert’, *pseudowetenschappelijk* ‘pseudoscientific’ (*WNT*)², Swedish *pseudopolitiker* ‘fake politician’ (*SAOB*)³). However, these developments did not take place in all languages at the same time. For instance, where *pseudo-* has a long history in French, Swedish words beginning with *pseudo-* appear relatively late. The first attestation of *pseudo-* in French (*TLFi*⁴ and *OED* (s.v. *pseudo-*)) is an isolated word in Old French (*pseudoprophete* ‘pseudoprophet’), adapted from post-classical Latin. It appears again in late Middle French (*pseudo-catholique* ‘pseudo-catholic’) and in the 17th century in adaptations from Latin (e.g. *pseudo-médecin* ‘pseudo-doctor’). Native formations in French are found from the 19th century onwards. The vitality of the morpheme is nowadays considerable, both in everyday and scientific language, in particular to construct nouns in biochemistry, chemistry and pathology (e.g. *pseudo-acide* ‘pseudo-acid’, *pseudo-lipome* ‘pseudo-lipoma’). In Swedish, by contrast, the earliest example is *pseudonym* from the beginning of the 18th century (*SAOB*). Like in the other languages, many early examples are from the scientific domain (e.g. *pseudohexagonal*, *pseudoskop* ‘pseudoscope’). Interestingly, in Swedish *pseudo* has been in use as an independent word for over a hundred years, e.g. by the author August Strindberg: *Stilen var förvänd och underskriften pseudo* ‘The style of writing was perverted and the signature was fake’ (1877).

From what precedes, it is clear that *pseudo-* is still productive as a left constituent in neoclassical compounds in modern languages, such as English

² *Woordenboek der Nederlandsche Taal*: <https://gtb.ivdnt.org/search/?owner=wnt>

³ *Svenska Akademiens Ordbok*: <https://svenska.se/>

⁴ *Trésor de la langue française informatisé*: <http://stella.atilf.fr/>

pseudomonas or *pseudoponium*. Neoclassical compounds, also called ‘internationalisms’, are mostly used to coin new technical or scientific terms (Bauer, 1983: 213-216; Iacobini, 2015: 1662). From the domain of neoclassical compounds, *pseudo-* came to be extended to formations with non-classical second constituents, as in *pseudo-Christian* or *pseudocolor*. The morphological status of *pseudo-* and other elements of classical origin has been the subject of extensive debate. In the original neoclassical compounds, *pseudo-* functions as an ‘Initial Combining Form’ (ICF) which can form a compound with a, likewise neoclassical, ‘Final Combining Form’ (FCF), as in *pseudonym*. Like prefixes, combining forms are bound morphemes that do not appear as independent words (at least not initially), but they differ from prefixes such as *super-* in that they tend to have higher ‘semantic density’, and in that they almost always end in a vowel (usually *-o*). Furthermore, prefixes like *super-* cannot combine with FCFs (Bauer, 1983: 214-215). However, the morphological status of *pseudo-* changes as soon as it starts combining with free lexemes, as in *pseudoscience* or *pseudo-British*. In such formations, *pseudo-* behaves more like a prefix, according to four criteria of prefixhood outlined in Prčić (2005: 329): it forms an endocentric relation with the right-hand member in a complex word, it has modifying meaning, it has a fixed position as a left-hand member, and its productivity increases. In our view, however, that does not make *pseudo-* a full-fledged prefix, as Prčić suggests (he categorizes it as a ‘qualifying prefix’), because it does not meet some of the other properties he uses to distinguish ICFs from prefixes. For example, the criterion that prefixes cannot combine with FCFs is not met; on the contrary, *pseudo-* is still best analysed as an ICF in neoclassical compounds (compare *pseudophobia* vs. **cophobia*). Furthermore, some of Prčić’s criteria, like semantic substance or productivity, are gradual rather than absolute, which makes them less convincing as indications of prefixhood (Kastovsky, 2009, for further discussion). Following Iacobini (2015), we therefore consider *pseudo-* a ‘neo-classical combining form’, which reflects its special status in morphology as a bound morpheme of Ancient Greek origin. A neoclassical combining form is best considered as a

morphological type in its own right, even though it has properties in common with both compounding and derivation.⁵

2.2 Morphological properties

As described above, Greek *pseudo-* was borrowed into a variety of European languages. Initially, Greek or Latin formations were borrowed as a whole (e.g. English *pseudonym*), but later *pseudo-* came to be combined with native words in a wide set of European languages, as shown in the contemporary examples (1-4) from the TenTen web corpora (see Section 4.2 for more information on these corpora).

- (1) [IT] Il fatto è che il calcio [...] è diventato in Italiai [sic] uno sport per vecchi ricchi e per *pseudotifosi* in pantofole.

“The thing is that, in Italy, football has become a sport for rich old people and pseudo-supporters in slippers”

- (2) [ENG] This must be a new trend in *pseudo left* thinking, a total failure to understand basic logic.

- (3) [DU] Een *pseudo historische roman* gebaseerd op oeroude Britse bronnen

“A pseudo historical novel based on age-old British sources”

- (4) [DA] Det, der er ‘*pseudo*’ her, er de ‘*problemer*’ politikerne søger at løse.

“What is ‘pseudo’ here, are the ‘problems’ politicians claim to be solving.”

⁵ An alternative term is ‘confix’, which, together with affixes, is sometimes subsumed under the term ‘combimeme’ to refer to any bound form that is not a lexeme. Since, however, ‘confix’ can mean different things in different philological traditions (see Giannouloupoulou, 2006 and references there), we do not use it in this paper.

At the morphological level, *pseudo-* combines with both nouns (1) and adjectives (2). Of particular interest are constructions where *pseudo-* takes scope over a noun phrase (3), or where *pseudo* is developing into an independent adjective meaning ‘fake’, as in (4). The latter process is known as ‘debonding’, i.e. a type of degrammaticalisation defined by Norde (2009: 186) as “a composite change whereby a bound morpheme in a specific linguistic context becomes a free morpheme”. Debonding has been shown to be a relatively widespread phenomenon in evaluative morphology (e.g. Van Goethem and De Smet, 2014; Norde and Van Goethem, 2018; Van Goethem and Norde, 2020). We will discuss it more extensively in Section 3.2.

2.3 Semantic profile

According to the *OED*, *pseudo-* forms nouns and adjectives with the meaning ‘false, pretended, counterfeit, spurious, sham; apparently but not really, falsely or erroneously called or represented, falsely, spuriously’. It conveys an approximative meaning (‘close to X’) either in a classifying (privative) sense (‘close to X but not X’), as in example (5) where *pseudogene* is explicitly distinguished from *gene*, or in a qualifying (evaluative) sense (‘close to X but a bad instance of X’), as illustrated in (6).

- (5) [ENG] The main obstacle in molecular diagnosis of CAH is amplification of *pseudogene* during polymerase chain reaction of CYP21A2. All attempts focus on discrimination of *pseudogene* from *gene* [...]
- (6) [GE] der Film sei schamlos kindisch, *pseudoreligiös* und unaufrichtig und die teuerste Hollywood-Komödie aller Zeiten
“the movie is shamelessly childish, pseudo-religious and insincere and the most expensive Hollywood comedy of all times”

Generally speaking, the classifying sense is mostly found in neoclassical compounds, belonging to technical (scientific) language (e.g. *pseudomonas*), but it also occurs in native formations (e.g. Dutch *pseudokroep* ‘pseudo croup’). The evaluative meaning is typical of contemporary native formations (non-technical language), as illustrated in (1-4), which indeed show that *pseudo-* is expanding to collocational contexts in which it is typically (but not necessarily) used with an ironic tone or negative connotation.⁶

Because of the approximative meaning *pseudo-* may convey, we consider it a case of evaluative morphology. More specifically, *pseudo-* expresses approximation in a classifying (privative) or qualifying sense (see Masini *et al.*, 2023 for a state of the art on approximative morphology). The field of evaluative morphology has been studied extensively in previous years (e.g. Bauer, 1997; Körtvélyessy, 2015, and the papers in Grandi and Körtvélyessy, 2015), but the focus of attention has been largely restricted to the subdomains of diminution and augmentation (including intensification) (e.g. Dressler and Merlini Barbaresi, 1994; Jurafsky, 1996; Schneider, 2003; Bakema and Geeraerts, 2004; Prieto, 2005; Körtvélyessy and Štekauer, 2011; Efthymiou, 2015; Rainer, 2015). With the notable exception of diminutive markers being used as attenuation strategies or derived from approximative values (cf. Dressler and Merlini Barbaresi, 1994; Merlini Barbaresi, 2015; Grandi, 2017), we still know very little about how approximation works within morphology, even less so from a contrastive perspective.

The most studied approximative morpheme is English *-ish*, which has undergone remarkable semantic and morphological developments in the course of time. Approximative *-ish* now expresses vagueness and speaker attitude, while it has been expanding its host classes from adjectives (e.g. *warmish*) to a number of other categories (e.g. *9-ish*, *okay-ish*) (Oltra-Massuet, 2017; Kempf and Eitelmann, 2018; Eitelmann, Haugland and Haumann, 2020; Eitelmann and Haumann 2023). It has also been discussed as an example of

⁶ As mentioned above, a more elaborate semantic analysis falls beyond the scope of this paper, but we reserve it for a separate study (see also Vassiliadou *et al.*, 2023 on the semantics of *pseudo-* in Greek and French).

degrammaticalization because of its recent autonomous or ‘debonded’ uses (Kuzmack, 2007; Norde, 2009), as in (7) (see further Section 3.2).

- (7) Ali: So, you feeling any better yet? - Me: Eh, *ish ish*
(urbandictionary.com)

Two other recent studies focused on the emergent approximative use of Italian *simil-* (e.g. *simil-marsupio* ‘sort of marsupium/pouch’, *freddo simil siberiano* ‘Siberian-like cold’) (Masini and Micheli, 2020) and the productivity, semantic profiles and categorical flexibility of a series of Dutch morphemes with ‘fake’ semantics (e.g. *kunstgras* ‘artificial grass’, *namaak-wasabi* ‘fake wasabi’, *neppe cupcake* ‘fake cupcake’) (Van Goethem and Norde, 2020). The present study aims to contribute to this growing body of research from a typological perspective, by focusing on a single approximative morpheme that has been borrowed into various languages. It is to this typological framework that we now turn.

3 Typological framework

3.1 The Germanic and Romance Sandwiches

Typological research has shown that languages belonging to the same (sub-)family may change at different paces, resulting in, among others, cross-linguistic differences with respect to the degree of analyticity. Lamiroy (2011: 167), for instance, argues that French and English are generally more analytical than Spanish and German, respectively. A higher degree of analyticity correlates with a lower degree of inflection, and vice versa. Thus, English and French are highly analytical and show little inflection, whereas Spanish and German are more synthetic and show much richer inflection. Based on empirical evidence from different linguistic subfields, two clines have been proposed for the Germanic and Romance languages, i.e. the so-called ‘Germanic Sandwich’ and the – less-known

– ‘Romance Sandwich’ clines (see, e.g., van Haeringen, 1956; Hüning *et al.*, 2006; Lamiroy, 2011; König and Gast, 2018).

According to the so-called ‘Germanic Sandwich’ hypothesis, Dutch is not only geographically, but also linguistically ‘sandwiched’ between English and German. Along this cline, the continental Scandinavian languages (SCAND in Figure 1) would be situated in-between English and Dutch, and Icelandic (ICEL in Figure 1) would be situated at the most synthetic endpoint of the Germanic cline (Lamiroy, 2011: 170).

Within the Romance languages, French is assumed to be the most analytical language and Romanian the most synthetic one. In-between we find Italian, Spanish and Portuguese respectively (Lamiroy, 2011: 170).

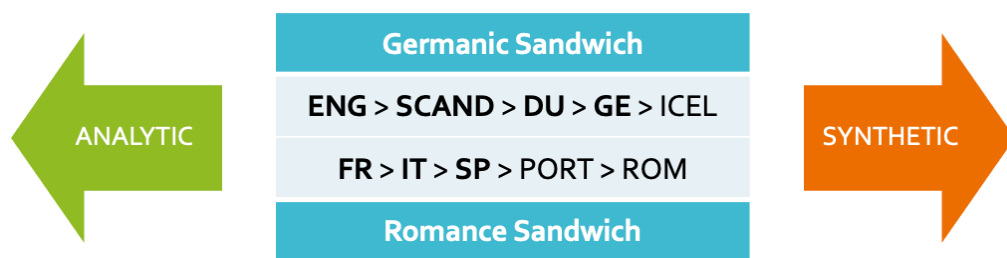


Figure 1. The Germanic and Romance Sandwiches (based on Lamiroy, 2011: 170; languages in bold form part of our dataset)

As mentioned previously, these clines can be found in several linguistic domains. Lamiroy (2011) provides data from three Romance (French, Italian, Spanish) and three Germanic (English, Dutch, German) languages, covering key properties in morphology and syntax, such as adjectival and verbal inflection and the use of auxiliaries. In order to account for these clines, she assumes that internal and external factors interact in a cumulative way.

3.2 Morphological cohesion and debonding in Germanic and Romance

Obviously, the degree of analyticity and weight of inflection may have an impact on the bound or free use of morphemes and the possible development of bound into free morphemes, which is labelled ‘debonding’. This has been shown in previous research by Van Goethem and De Smet (2014) who identify three factors that facilitate debonding of bound morphemes and subsequent reanalysis into an adjective in French, English and Dutch:

- the qualifying semantics of the morpheme subject to debonding;
- the low prosodic and morphological cohesion of the sequence;
- the absence of adjectival inflection.

The impact of these three factors can be illustrated by formations with English *key*, French *-clé* and Dutch *sleutel-* (e.g. *key role*, *role-clé*, *sleutelrol*). In these sequences, English *key* and its French and Dutch equivalents *clé* and *sleutel* have developed the qualifying meaning ‘very important, crucial’. When a lexeme develops a specialized meaning when embedded in a compound, often a more abstract evaluative meaning, it is commonly called an affixoid (Booij, 2010: 57). However, the affixoid may undergo debonding if certain conditions are met. In the three languages, the item has developed evaluative semantics and is semantically close to a qualifying adjective (‘very important, crucial’), which is in favour of N>A reanalysis. Nevertheless, Dutch *sleutel-* does not debond (**een heel sleutele rol* ‘lit. a very key role’), whereas English *key* and French *clé* do occur in debonded sequences (e.g., *a very key role* / *un rôle vraiment clé*) (see also Amiot and Van Goethem, 2012, for a detailed comparison of Dutch and French, and De Smet, 2012, for diachronic support in favour of the N-to-A reanalysis in English). This shows that the strong prosodic and morphological cohesion typical of Dutch word formation, combined with a more elaborate inflectional system compared to (spoken) French and especially English, may prevent the item from debonding, despite its qualifying semantics (Van Goethem and De Smet, 2014).

Debonding has been shown to be widespread in evaluative morphology, affecting different types of bound morphemes, including affixoids and affixes

(among others Norde, 2009; Norde and Van Goethem, 2015; Oltra-Massuet, 2017). Accordingly, we would expect (qualifying) *pseudo-* to debond as well, and more decidedly so in languages towards the ‘analytic’ pole of the ‘sandwiches’, especially in English and French. After all, these languages favour analytical structures above synthetic ones and show less elaborate inflection (which potentially prevents a debonding process).

Besides these typological factors, Van Goethem and Norde (2020) show that debonding and – possibly – subsequent adjectival reanalysis also correlate with the productivity of the morpheme within a particular language. More specifically, their analyses of eight Dutch ‘fake’ morphemes suggest that their debonding and categorical flexibility is triggered by an interplay of type frequency and semantic coherence, two factors that are inversely correlated (Barðdal, 2008).⁷

These observations imply that the morphological properties and – specifically – the debonding of *pseudo-* is determined not only by typological factors (such as cross-linguistic differences in morphological cohesion and degree of inflection), but also by language-internal factors such as the productivity and semantic coherence of the formations.

4 A cross-linguistic corpus study of *pseudo-*

4.1 Research questions

Based on previous research into Italian and Dutch ‘approximative’ morphemes (Masini and Micheli, 2020; Van Goethem and Norde, 2020) and into ‘morphological cohesion’ and ‘debonding’ in English, French and Dutch (Van

⁷ Specifically, the eight Dutch morphemes in Van Goethem and Norde (2020) could be ranged on a cline where the morphemes belonging to the upper part of the cline (*nep-* ‘fake, false’, *fake-*, *namaak-* ‘imitation’, *imitatie* ‘imitation’) undergo debonding more easily, due to low degree of semantic coherence of the formations and high type frequency. Conversely, morphemes belonging to the lower part of the cline (*schijn-* ‘apparent’, *fop-* ‘lit. fool’, *lok-* ‘lit. lure’, *kunst-* ‘artificial’) which are characterized by strong semantic coherence (i.e. they form lexicalized compounds) and low type frequency, are not subject to debonding or further adjectival reanalysis.

Goethem and De Smet, 2014, among others), we formulate the following two research questions:

1. Do our findings with respect to the morphological properties, frequency and productivity of *pseudo-* confirm the clines from synthetic to analytic languages, as observed for the Germanic and Romance languages (see Section 3.1)?
2. Using linear regression modelling (Levshina, 2015), is there a correlation between the morphological productivity of *pseudo-* in a particular language and its degree of debonding (see Section 3.2)?

4.2 Data and methods

In order to gain a better understanding of the morphological behaviour and productivity of *pseudo-* words in contemporary European languages, we carried out a cross-linguistic corpus analysis based on 1000-token samples in each of the eight languages under study (Danish, Dutch, English, French, German Italian, Spanish, Swedish).

The samples were extracted from the TenTen web corpora (Kilgarriff *et al.*, 2014) available at the SketchEngine platform. The TenTen corpora are very large corpora consisting of random sentences crawled from the web that lend themselves well to comparative research (see, for instance, Van Goethem and Koutsoukos, 2022, on denominal verb formations in Germanic languages). Although these are webcorpora, they contain both formal and informal genres, and both technical and non-technical texts. We used regular expressions to capture *pseudo* as a free morpheme or strings starting with *pseudo*, e.g. `[word="pseudo|pseudo.+ "].`⁸ For each language, we downloaded a random sample of 2000 occurrences, which were manually annotated in Excel until we obtained a set of 1000 relevant examples per language.

⁸ For some languages, we used slightly adapted regular expressions, e.g. `[word="(?)pseudo|(?)pseudo.+"]` for German, in order to include all nouns (which are usually written with an upper case initial), or `([word="pseud|pseud.+"]|[word="seud|seud.+"])`, in order to include orthographical variants and to accommodate possible inflection.

All relevant occurrences were annotated for the following orthographic and morphological properties:

- (Orthographic) bonding (one word, hyphen, two words);
- Construction type (neoclassical compound, native formation, clipping);
- Reference lemma (R1), i.e. the right-hand member (with bound *pseudo-*) or the word or phrase (with debonded *pseudo-*) modified by *pseudo(-)*;
- Scope of *pseudo(-)* (N, A, NP, AP, V, Adv, Pron) for native formation;
- Presence of inflection on debonded *pseudo*.

We also analysed word frequency and the productivity of native formations (thus excluding neoclassical compounds), based on ‘type/token ratio’ (TTR), ‘potential productivity’ (PP) and ‘global productivity’ (GP) (Baayen and Lieber, 1991; Baayen, 2009). In the following, we illustrate the different levels of annotation.

First, **bonding** was used as a merely orthographic criterion; we distinguished between sequences spelled as one word (8), with a hyphen (9) or as two separate words (10).

- (8) [ENG] Parapsychology is a *pseudoscience* concerned with the investigation of paranormal and psychic phenomena which includes telepathy, precognition, (...)
- (9) [DU] Verder maken aanhangers van *pseudo-wetenschappen* vaak gebruik van wetenschappelijk jargon, zonder dat evenwel duidelijk wordt gemaakt wat daar precies mee bedoeld wordt
“Furthermore, proponents of pseudo-sciences often use scientific jargon without making clear what exactly is meant by it”
- (10) [ENG] I've never been a fan of *pseudo psychology* and the left brain / right brain rubbish that is out there drives me mad.

With respect to **construction type**, we annotated whether the sequence formed a neoclassical compound (11) or a native formation (12), namely a formation with a synchronically recognizable and independent base attaching to *pseudo-* (see

uomini in (12)). A third category was assigned to clippings, which refers to the deletion of one or more syllables from multisyllabic words (Blank, 2001:1604, 1605), e.g. French *pseudo* as a clipped form of *pseudonyme* (13).

- (11) [ENG] Then, it was reported that the venom affects the organization of the cellular cytoskeleton and *pseudopodia* formation of epithelial cells.
- (12) [IT] ho assistito all'abbandono di ammalati da parte di *pseudo-uomini*
“I witnessed the abandonment of sick people by pseudo-men”
- (13) [FR] Choisir le bon *pseudo* : Evitez les *pseudos* avec plusieurs chiffres et optez pour un *pseudo* original et qui saura mettre votre personnalité en avant : Le nom de votre héros ou groupe musical préféré, par exemple.
“Choose the right nickname: Avoid nicknames with several numbers and choose an original nickname that will highlight your personality: The name of your favourite hero or musical group, for example”

In the native formations, we annotated the **scope of *pseudo-*** (N/NP, A/AP, V/VP, Adv, Pron). Examples (14) to (18) illustrate these different categories.

- (14) [SP] Desgraciadamente los mexicanos les encanta indignarse por *pseudoartistas* cuando el país esta hundido en la mierda... [Noun]
“Unfortunately, Mexicans love to get indignant about pseudoartists when the country is in deep shit...”
- (15) [FR] Dans cette partie de la ville la plupart des maisons sont récentes et construites dans un style *pseudo russe ancien*. [Adjectival Phrase]
“In this part of the city most of the houses are new and built in a pseudo old Russian style.”
- (16) [ENG] IF we ever have credible, verifiable reports of actual McCain staffers making remarks such as have been *pseudoreported* by Mass

Media Podpeople, THEN it'd be time to get out the tar and feathers
[Verb]

(17) [ENG] Based on this information, following activation of its PtS function, a radioterminal may randomly, *pseudo-randomly* and/or in a predetermined way, pick a channel and send an initial message to the base station over that channel. [Adverb]

(18) [GE] Ein ziemlich lächerlicher Titel für einen Studiengang, da *es* sich nicht nur nach Möchtegern und *Pseudo* anhört, sondern auch eindeutig kein Zertifikat impliziert. [Pronoun]⁹

“A somewhat ridiculous title for a study programme, because it not only sounds like Wannabe and Pseudo, but also clearly implies that no certificate will be issued.”

Finally, we annotated the presence of **inflection** on debonded *pseudo*, like the number inflection in the French example in (19).

(19) [FR] Pour terminer, je te le répète, quelle chance que mes enfants, n'ont pas à connaitre des *pseudos* ‘instits’ de ton genre !

“Finally, I repeat, how lucky my children are that they don't have to know pseudo ‘teachers’ like you!”

In the next sections, we report our findings on the cross-linguistic behaviour of *pseudo-* with respect to its construction types (4.3.1), word frequencies (4.3.2) and productivity (4.3.3). In Section 5, we analyse the potential correlation between debonding and productivity.

⁹ Overall, scope over pronouns is very rare in our corpus, but in this particular example we annotated *es* ‘it’ as the R1 because it is the subject of the sentence, with *pseudo* in predicative position.

4.3 Results

4.3.1 Construction types

Figures 2 and 3 show the token frequencies of each construction type (including neoclassical compounds) and of each POS (native formations only).

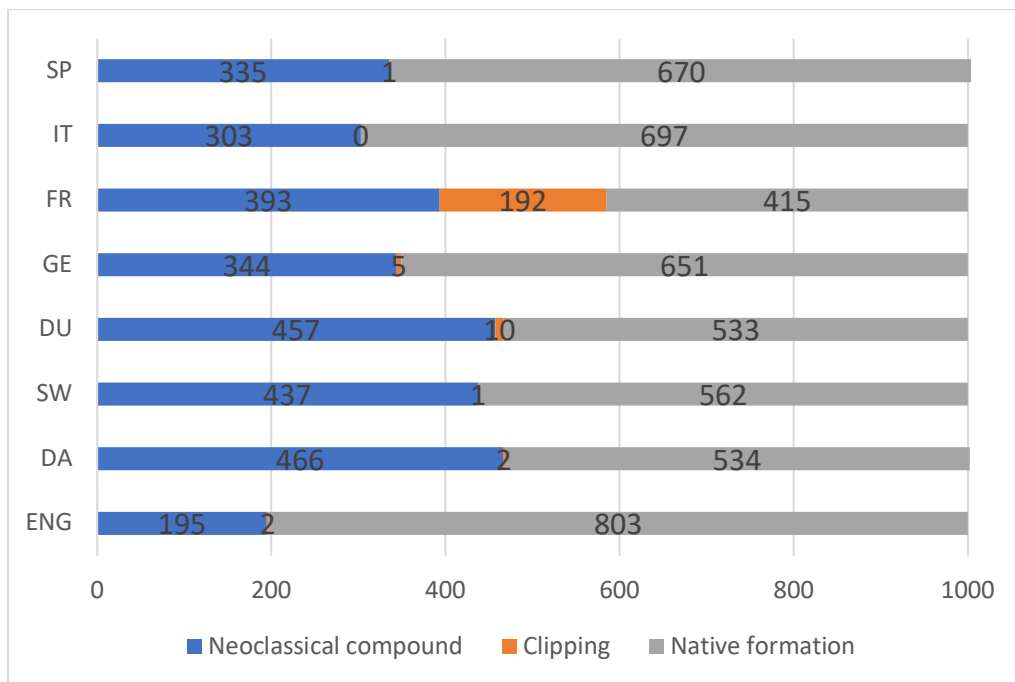


Figure 2. Construction types (based on token frequencies)

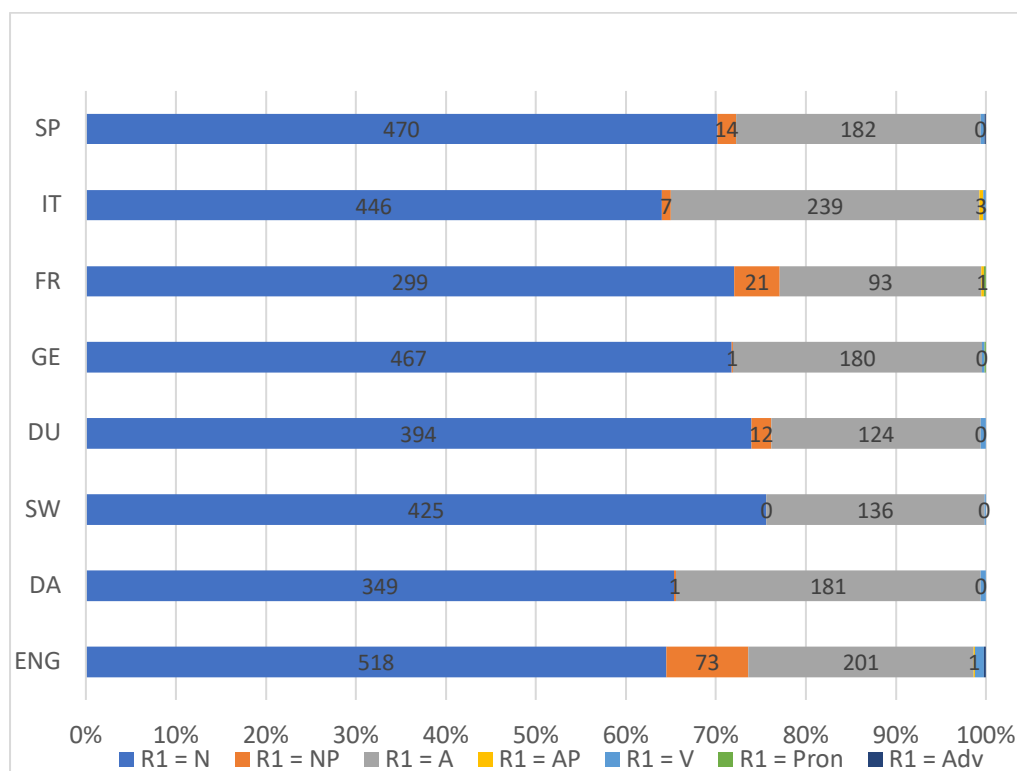


Figure 3. POS of R1 for native formation only (based on token frequency)

In general, we notice a slightly higher proportion of neoclassical compounds in Germanic than in Romance (except for English), and a remarkably high proportion (19.2%) of clippings in French (almost exclusively *un pseudo* as short for *un pseudonym* ‘a nickname’). Figure 3 furthermore indicates that *pseudo-* combines more often with nominal R1s than with adjectival ones; the other parts of speech occur only rarely. Italian shows the highest proportion of adjectival bases. English has the most divergent construction type profile, showing most variation in the R1 position (e.g. V, ADV); it also shows the highest proportion of cases in which *pseudo-* has scope over an NP (9.07% of its native formations). Swedish, Danish and (to a lesser extent) Dutch show quite similar construction type profiles.

4.3.2 Word frequencies

Table 1 gives an overview of the most frequent types *pseudo-* combines with in the different languages in our dataset (all lexemes with a token frequency ≥ 10). It shows that *pseudonym* is the most frequent *pseudo-* type in all eight languages under study. Its token frequency ranges from 182 in the English sample to 456 in the Danish one.

Table 1. Lexemes attaching to *pseudo-* with a token frequency ≥ 10

Rank	English	Danish	Swedish	Dutch	German	French	Italian	Spanish
1	<i>pseudonym</i> (182)	<i>pseudonym</i> 'pseudonym' (456)	<i>pseudonym</i> 'pseudonym' (434)	<i>pseudoniem</i> 'pseudonym' (442)	<i>Pseudonym</i> 'pseudonym' (330)	<i>pseudonyme</i> 'pseudonym' (384)	<i>pseudonimo</i> 'pseudonym' (297)	(p)seudónimo 'pseudonym' (311)
2	<i>science</i> (62)	<i>videnskabelig</i> 'scientific' (65)	<i>vetenskap</i> 'science' (137)	<i>wetenschappelijk</i> 'scientific' (42)	<i>wissenschaftlich</i> 'scientific' (40)	<i>science</i> 'science' (24)	<i>scientifico</i> 'scientific' (35)	<i>ciencia</i> 'science' (59)
3	<i>gene</i> (21)	<i>videnskab</i> 'science' (52)	<i>vetenskaplig</i> 'scientific' (57)	<i>wetenschap</i> 'science' (40)	<i>Wissenschaft</i> 'science' (15)	<i>scientifique</i> 'scientific' (22)	<i>scienza</i> 'science' (26)	<i>cientifico</i> 'scientific' (38)
4	<i>random</i> (21)	<i>tvilling</i> 'twin' (20)	<i>debatt</i> 'debate' (26)	<i>kroep</i> 'croup' (12)	<i>Krupp</i> 'croup' (12)	<i>scorpion</i> 'scorpion' (15)	<i>artrosi</i> 'arthrosis' (11)	<i>aleatorio</i> 'random' (13)
5	<i>scientific</i> (21)	<i>intellektuel</i> 'intellectual' (10)		<i>eindheffing</i> 'final tax' (11)	<i>religiös</i> 'religious' (10)			<i>religioso</i> 'religious' (10)
6	<i>code</i> (20)			<i>graan</i> 'grain' (11)				<i>intelectual</i> 'intellectual' (10)
7	<i>ephedrine</i> (14)			<i>wetenschapper</i> 'scientist' (11)				
8	<i>christian</i> (11)			<i>ephedrine</i> 'ephedrine' (10)				
9	<i>aneurysm</i> (10)							
10	<i>religious</i> (10)							

The languages also show a comparable distribution of the lexemes: *pseudonym* stands out as the most frequent type and is generally followed by far less frequent types referring to 'science', 'scientific' or 'scientist', or to scientific terminology (e.g. *aneurysm* [ENG], *artrosi* 'arthrosis' [IT], *Krupp* 'croup' [GE], *ephedrine* 'ephedrine' [DU]), and by items meaning 'intellectual', 'religious' or 'random'.

This distribution is illustrated in greater detail for German and Italian in Figures 4 and 5 (for the other languages, the graphs look very similar).

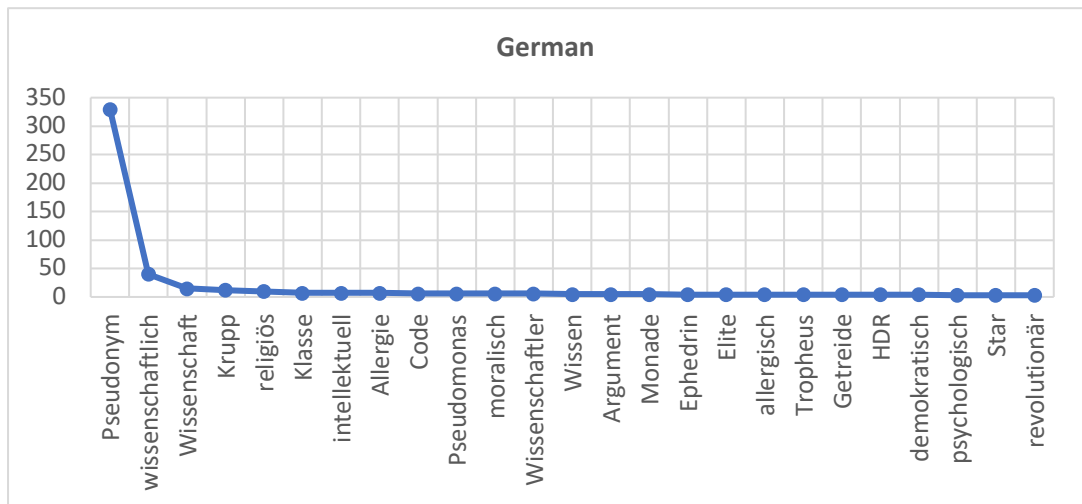


Figure 4. Word frequency of *pseudo-* words in German

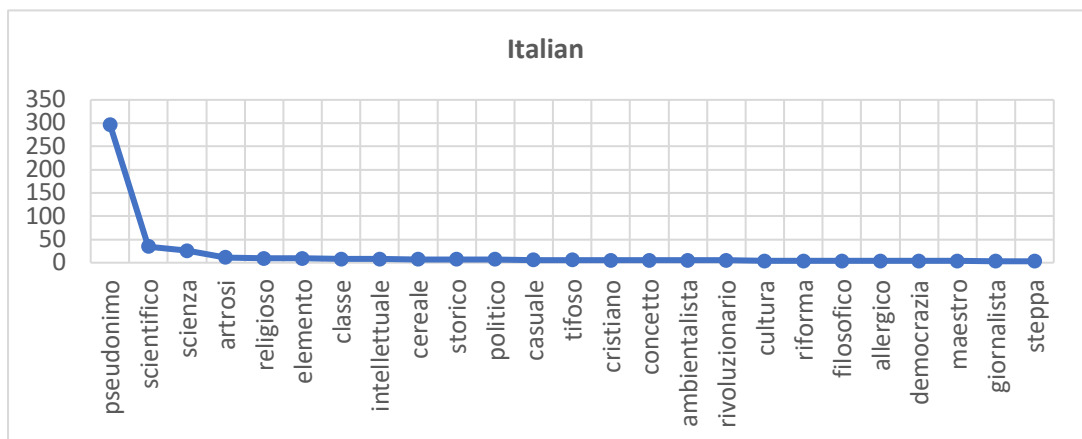


Figure 5. Word frequency of *pseudo-* words in Italian

4.3.3 Productivity

Figure 6 compares the productivity of *pseudo-* in native formations across all parts of speech. Neoclassical compounds, such as *pseudonym* or *pseudomonas*, and clippings are excluded from this graph. The scores indicate type/token ratio (TTR) and potential productivity (PP), the latter referring to the number of hapax legomena divided by the number of tokens (Baayen, 2009).

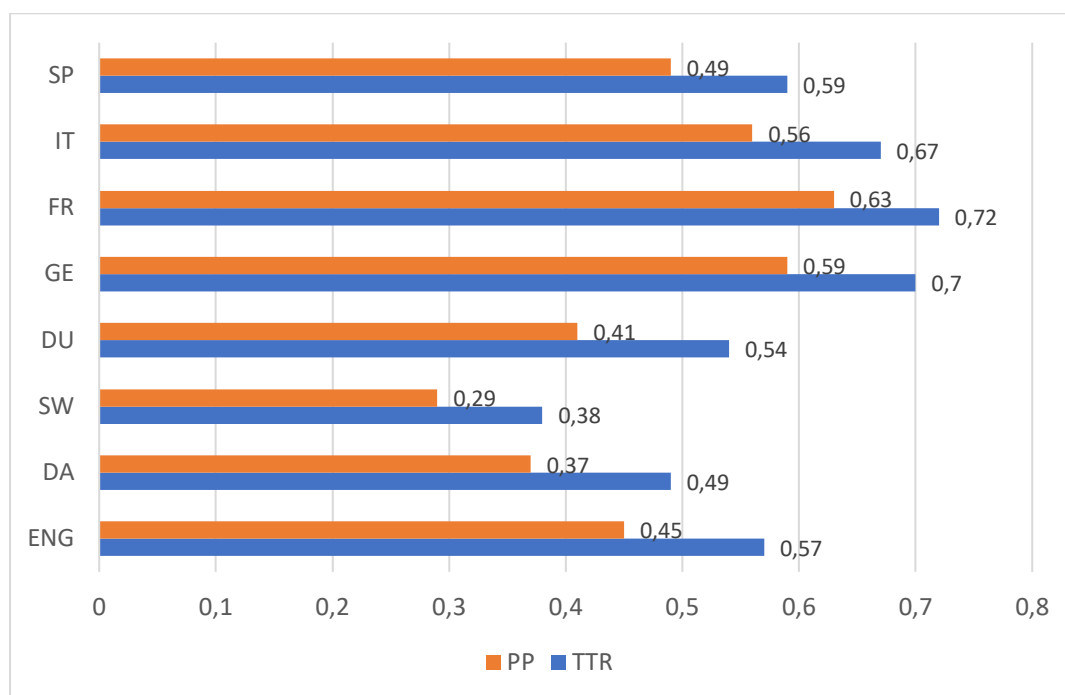


Figure 6. Productivity of the *pseudo-* words (in native formation, all POS)

In sum, we observe an overall higher productivity of *pseudo-* in Romance than in Germanic, with the highest productivity scores found in the French sample. In the Germanic languages, German stands out with the highest productivity scores, Danish and Swedish with the lowest scores. The lower scores for Dutch and English may be due to competition with numerous native compounding elements and prefixes expressing similar meanings (see, for instance, Van Goethem and Norde, 2020 for Dutch or Cappelle *et al.*, 2023 for English), whereas Danish and Swedish often use adjectives such as *falsk* ‘fake’ instead of *pseudo-*constructions.

Given these differences, it can be useful to plot potential productivity against type frequency, a method proposed in Baayen and Lieber (1991: 818-819), which they term ‘global productivity’. Global productivity of *pseudo-* in the eight languages is shown in Figure 7, with potential productivity on the y-axis and type frequency on the x-axis. The plot distinguishes between the Romance and the Germanic languages (the former written in red characters, the latter in grey), and between nominal and adjectival bases (the former orange-dotted, the latter blue-dotted).

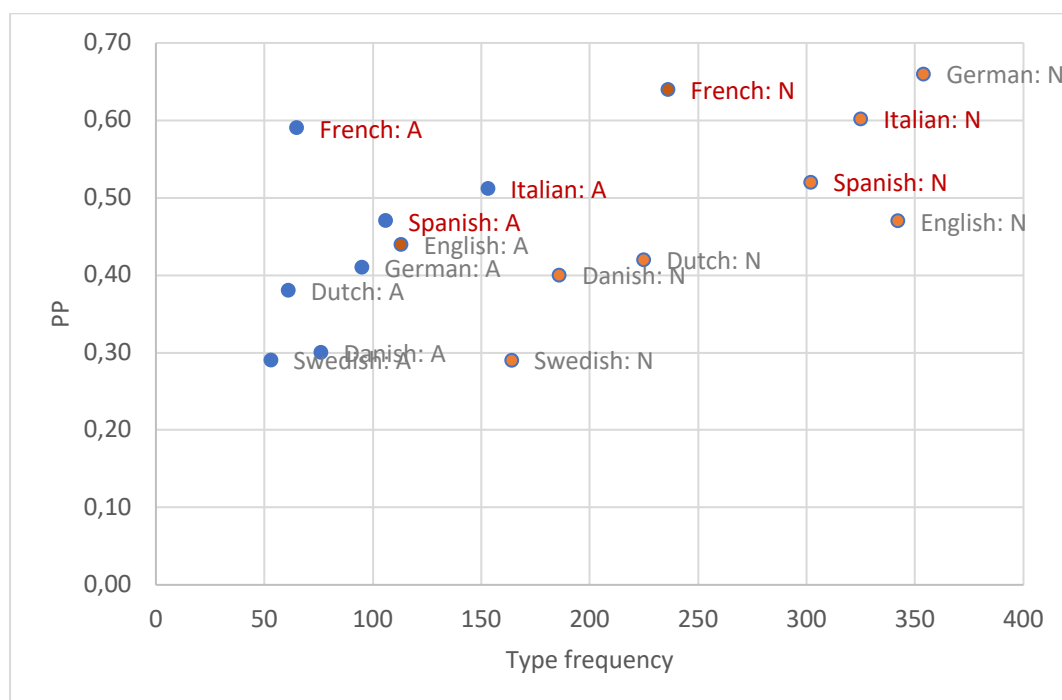


Figure 7. Global productivity (R1 = A or N)

Figure 7 clearly shows a higher global productivity of *pseudo-* when combined with nominal bases than with adjectival ones. Moreover, *pseudo-* is in general more productive in Romance than in Germanic. However, two exceptions to this tendency should be noted: when combined with nouns, *pseudo-* also has a high global productivity in English, and an even higher one in German. The group of German nouns combined with *pseudo-* even reaches the highest global productivity score; the difference with German adjectival bases is striking.

5 Degree of debonding and categorial flexibility

Having examined the construction types, frequency and productivity of *pseudo-* in Romance and Germanic, let us recall our second research aim, i.e. to investigate the debonding and categorial flexibility of *pseudo-* and identify which factors determine it. In 5.1 we analyze the degree of (de)bonding of *pseudo-*, defined here as a purely orthographic criterion, and in 5.2 we look into its categorial

flexibility, i.e. its potential to adopt adjectival or adverbial morpho-syntactic properties. In Section 5.3 we address the potential correlation between productivity and debonding.

5.1 Degree of (de)bonding

As we explained in 4.2, we operationalize debonding here as a purely orthographic criterion; it implies that the *pseudo-* sequences are written as two words. Sequences written as one word and hyphenated sequences have been considered instances of bound use. In figures 8a and 8b we show the degree of (de)bonding of the *pseudo-* words: Figure 8a includes the neoclassical compounds, which are always written as one word in our dataset; Figure 8b shows the distribution of the various types of bonding with the neoclassical compounds excluded.

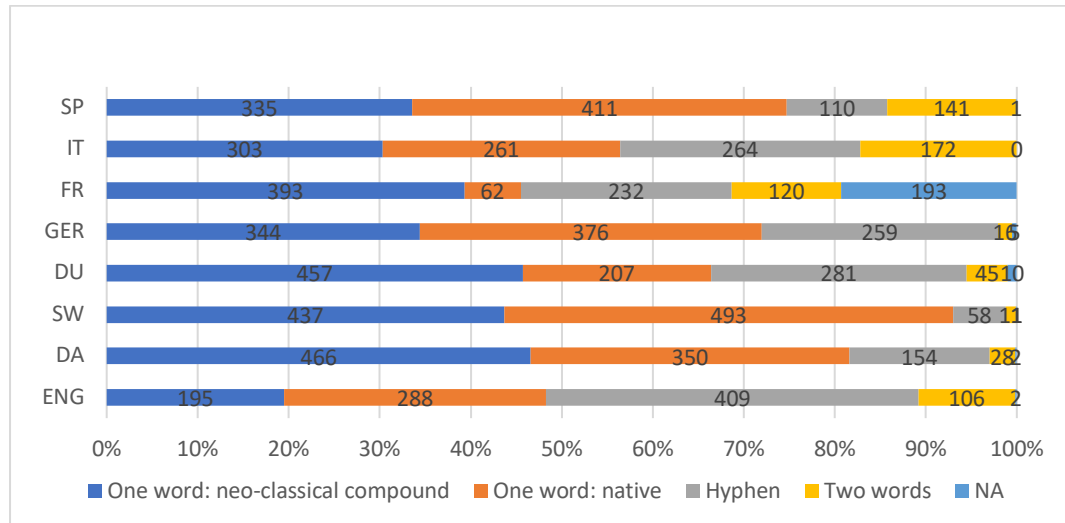


Figure 8a. Degree of (de)bonding (all tokens)

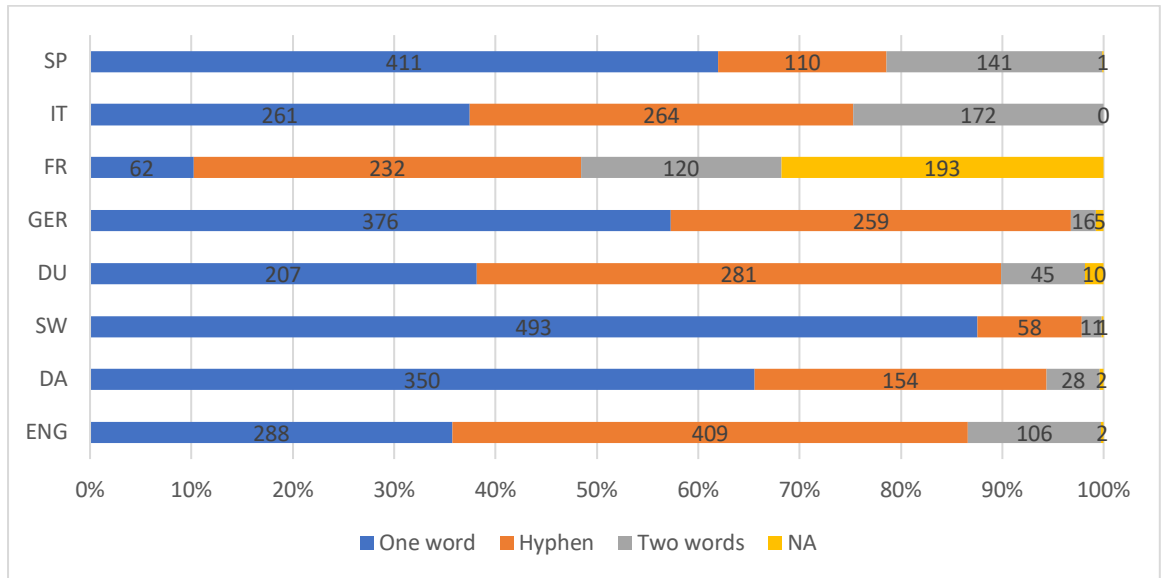


Figure 8b. Degree of (de)bonding (neoclassical compounds excluded)

As Figures 8a and 8b show, Romance languages show a higher proportion of spelling as two words overall, but in English, too, we find debonding in more than 10% of the data. The strongest tendency towards bonding, on the other hand, is found in Swedish. Compared to the latter Germanic language, Danish, Dutch, German and especially English feature more cases with hyphenation.

5.2 Debonding and categorical flexibility ratios

Within the cases of debonding, we find instances in which *pseudo-* displays ‘categorical flexibility’. This criterion refers to all morphological or syntactic features that flag adjectival or adverbial use. Such properties include scope over an NP, predicative use, coordination with an adjective, inflection, modification of a verb or adverb (the latter in case of adverbial use), as illustrated in (20)-(26).

- Scope over NP:

(20) [DU] In de sober versierde voorgevel vallen vooral de *pseudo dubbele voordeur* met gietijzeren raamdecoraties op

“In the soberly decorated facade, the pseudo double front door with cast iron window decorations stands out ”

- Predicative use:

(21) [SW] urbaniteten i fråga [...] blev först *pseudo*, sen sjangserade, och sen efterhärmdes

“The urbanity under discussion [...] first became pseudo, then faded and then it was being imitated”

- Coordination with an ADJ :

(22) [ENG] History cannot support two versions of the Macedonia story...one *pseudo and Slavic*, and the other one, mainstream and Hellenic [...]

- Inflection (in all, we found 15 cases of number inflection in French; 13 cases of number/gender inflection in Spanish):

(23) [FR] En fait, rien n'est plus faux que ces *pseudos appels écologistes* du type ‘un clic pour sauver un arbre’

“In fact, nothing is more false than these pseudo-environmentalist calls of the ‘one click to save a tree’ type”

(24) [SP] esas *seudas autobiografías* omiten la autocrítica y pecan de ‘yoísmo’.

“these pseudo-autobiographies omit self-criticism and err on the side of ‘I-ism’”

- Modification of V or ADJ (adverbial use of *pseudo*):

(25) [DU] Zeker als je ‘*pseudo samenwoont*’ en meer daar dan hier.

“Especially if you ‘pseudo live together’ and more there than here.”

(26) [FR] Cette histoire *pseudo policière* peut se lire indépendamment de la saga William Monk.

“This pseudo police story can be read independently of the William Monk saga.”

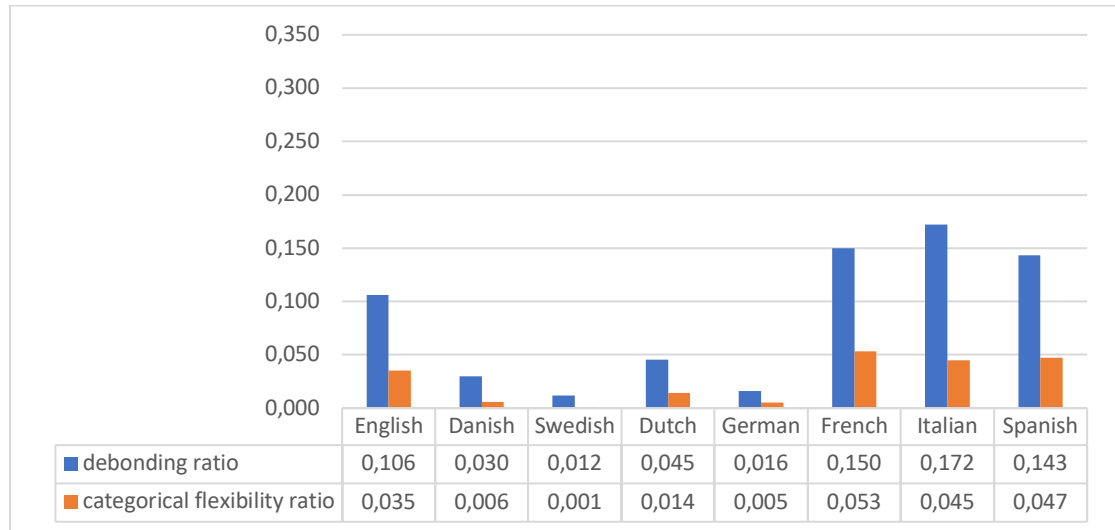


Figure 9a. Debonding and categorical flexibility ratios (all tokens)

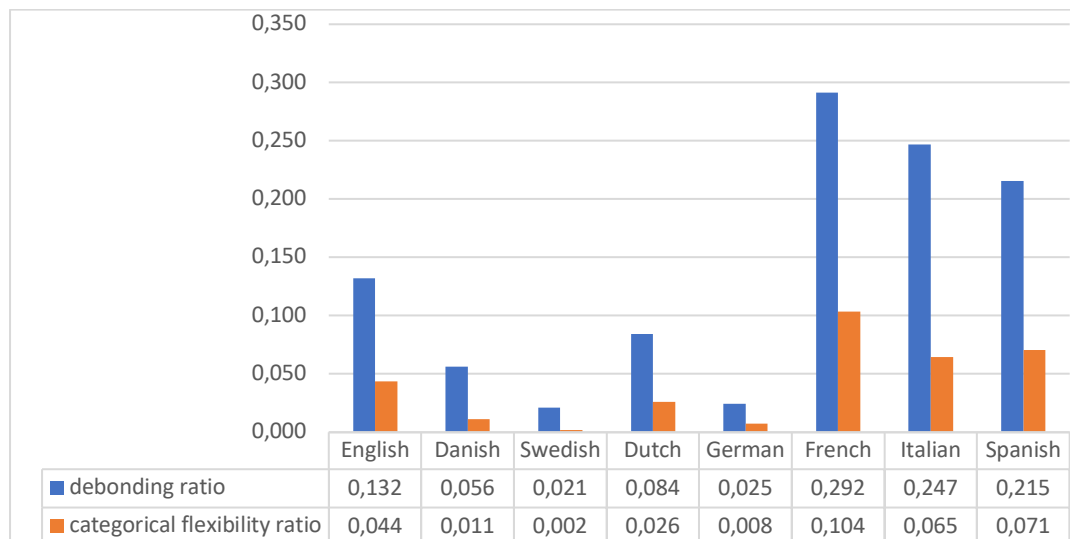


Figure 9b. Debonding and categorical flexibility ratios (neoclassical compounds excluded)

Figure 9a visualizes the ratio of debonding and categorical flexibility in the datasets of all eight languages. We obtained these proportions by dividing the number of instances of debonding or categorical flexibility by the number of tokens minus clipping constructions. First, we notice that Romance *pseudo-* debonds (graphically) much more than Germanic *pseudo-*. Second, it can be seen that Romance languages behave very much alike and display relatively high orthographic debonding and categorical flexibility. Third, within the Germanic languages, *pseudo-* debonds much more clearly in English than in the other Germanic languages, which have extremely low debonding and categorical flexibility ratios. However, since both the debonding ratio and the categorical flexibility ratio are affected by the number of neoclassical compounds in the dataset (neoclassical compounds never debond, which also implies they cannot be categorically flexible), we calculated the same ratios for the datasets excluding the neoclassical compounds. As shown in Figure 9b, this means, for instance, that French now has a higher debonding ratio than Italian and Spanish, and that the differences between English and the other Germanic languages are slightly smaller.

5.3 Correlation between productivity and debonding

The question now arises which factors may explain the different degrees of debonding we discussed in the preceding section. For Dutch ‘fake’ morphemes (see Section 3.2) Van Goethem and Norde (2020) found a statistical correlation between productivity and debonding. Since it emerges from our data that Romance *pseudo-* debonds more, while *pseudo-* is also more productive in Romance languages, we wanted to find out whether a similar correlation can be observed for *pseudo-* as well. For this analysis, we used the debonding ratio without the neoclassical compounds (cf. Figure 9b). To test whether there exists a statistically significant correlation between productivity (in terms of type/token ratio, see Section 4.3.3) and debonding (in terms of debonding ratio, see Section 5.2) we performed a correlation analysis with linear regression modelling (Levshina, 2015: 115-170), which is visualized in Figure 10.

As the incongruence of the data points and the regression line already suggest, the correlation is not significant ($p = 0.1003$). Figure 10 also shows why: a clear outlier is German, which has a high TTR but a low debonding ratio.

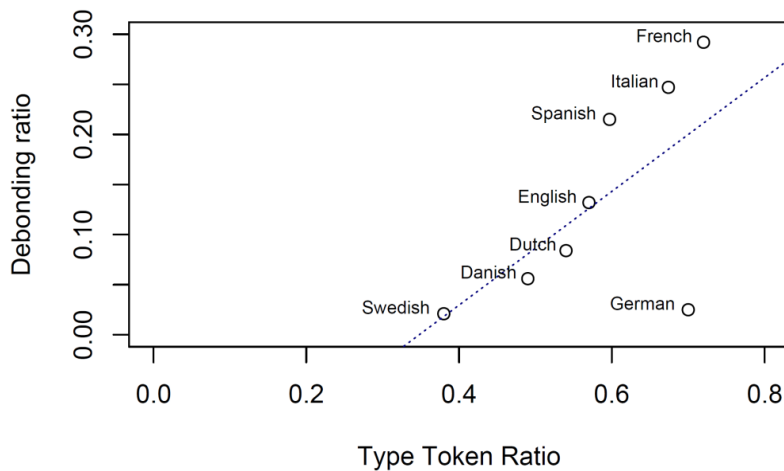


Figure 10. Correlation between TTR and debonding ratio (all languages)

For this reason, we performed a second regression analysis in which German was excluded.¹⁰ This time, we did find a statistically significant correlation between productivity and debonding ($p = 0.000832$ (***), Adjusted $R^2 = 0.893$, cor (Pearson's) = 0.954), as is also shown in Figure 11.

¹⁰ Opinions differ as to whether one can leave out data in order to obtain a significant effect, with some statisticians arguing that removing data points makes the data less representative of the population, whereas others find it is allowed for statistical reasons (see Levshina, 2015: 155 for brief discussion). We believe leaving out German is justifiable because our data are not meant to be representative of an entire population (which in this case would imply all languages featuring a *pseudo-morpheme*).

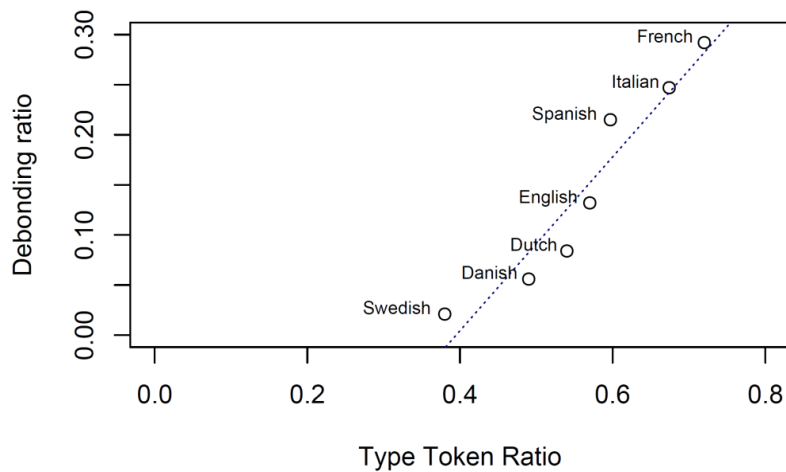


Figure 11. Correlation between TTR and debonding ratio (German excluded)

6 Conclusions and outlook

In this study, we analysed and compared the morphological and distributional properties of the combining form *pseudo-* within eight Germanic and Romance languages, including its debonding behaviour and categorical flexibility. Our contrastive study showed that *pseudo-* is a productive and vital morpheme, attaching mostly to nouns and – to a lesser extent – to adjectives, with variable properties in terms of productivity and debonding. The results of our investigation allow us to answer our two research questions.

The first research question was: does the behaviour of *pseudo-* in terms of morphological properties, frequency and productivity confirm the clines from synthetic to analytic known as Germanic and Romance Sandwiches? According to these generalizations, we would expect *pseudo-* to ‘debond’ more decidedly in languages towards the analytic pole of the sandwiches. Our prediction is indeed confirmed, given the high degree of debonding and categorical flexibility of *pseudo-* in English and French. However, language-specific factors also play a role (cf. also Van Goethem and De Smet, 2014). In languages with high morphological cohesion, like German, *pseudo-* tends to debond less, despite its

high productivity. This cannot be explained by orthographical rules alone, as compounds are also written as one word in Dutch, Danish and Swedish, yet we do find debonding in those languages (and more so than in German). On the other hand, the phonological resemblance of *pseudo* to Romance adjectives may play a role in reanalysing it as an adjective, boosting debonding. In French and Spanish, this even results in *pseudo-* sporadically adopting plural and/or gender inflection. The reanalysis may also have been favoured by the semantic shift that *pseudo-* has undergone from a purely privative meaning (typically associated with technical, scientific language, with a classifying function) to a more general evaluative meaning (typically associated with non-technical, common language), with possibly accounts for its higher productivity – a hypothesis to be tested in future research.

The second research question was: is there a correlation between morphological productivity and degree of debonding, as suggested in previous research (Van Goethem and Norde, 2020)? We did find a correlation, by using linear regression modelling: the languages where *pseudo-* is most productive are also those where *pseudo-* debonds more easily. However, the correlation is only found once we leave out German, which displays an inconsistent behaviour, as already observed.

Overall, our results tally with earlier research on Germanic and Romance Sandwiches and they also seem to be in line with De Smet's (2012: 607) observation that “language change has a sneaky quality”, namely that “language change often advances most easily where it is least obtrusive, apparently thriving on structural ambiguities and (possibly superficial) resemblances to existing patterns” (ibid.).

However, a number of issues remain to be addressed in follow-up studies. First and foremost, we need a more fine-grained analysis of the semantics of *pseudo-* constructions, especially in the light of the shift from a privative (classifying) semantics (cf. Cappelle *et al.*, 2023) to a more evaluative (qualifying) semantics, also attested in other cases (Masini and Micheli, 2020). The spread of the evaluative approximative meaning should be examined in more detail (also in relation to technical vs. non-technical language), together with its

impact on the productivity of *pseudo-* (and its debonding behaviour). In addition, a thorough intralinguistic comparison with other morphological markers of approximation would be highly desirable, to better understand their use and distribution. In particular, the high vs. low presence of competitors in specific languages may impact both productivity and debonding (cf. the wealth of ‘fake’ morphemes in Dutch, see Van Goethem and Norde, 2020).

In conclusion, we believe that our study, despite merely scratching the surface of *pseudo-*’s behaviour and development, paves the way to a new promising methodology of research into (evaluative) morphology that advocates for large-scale contrastive corpus studies. These allow to test generalizations but also to take into account language-specific properties that may have a role, doing justice to both.

Sources

Online Etymological Dictionary (Etymonline):

https://www.etymonline.com/word/pseudo-#etymonline_v_2781

Oxford English Dictionary (OED): <https://www-oed-com>

SketchEngine: <https://www.sketchengine.eu>; *TenTen* corpora available on SketchEngine used in this study:

- Danish: daTenTen17 (1.9 billion tokens)
- Dutch: nlTenTen14 (2.6 billion tokens)
- English: enTenTen15 (13.2 billion tokens)
- French: frTenTen17 (5.8 billion tokens)
- German: deTenTen13 (16.5 billion tokens)
- Italian: itTenTen16 (4.9 billion tokens)
- Spanish: esTenTen18 (16.9 billion tokens)
- Swedish: svTenTen14 (3.4 billion tokens)

Svenska Akademiens Ordbok (SAOB) (scientific dictionary published by the Swedish Academy): <https://svenska.se/>

Trésor de la langue française informatisé (TLFi): <http://stella.atilf.fr/>

Woordenboek der Nederlandsche Taal (WNT):

<https://gtb.ivdnt.org/search/?owner=wnt>

Acknowledgements

This paper was presented at the 2021 *Using Corpora in Contrastive and Translation Studies* conference (Bertinoro, Italy). We are grateful to the audience, as well as to the journal editors and the anonymous reviewers, for helpful comments and suggestions for improvement. Thanks are also due to Sarah Sippach and Benjamin Houitte (Humboldt-University of Berlin) for their assistance in annotating the German, Danish and Swedish data, and to Beatrice Bini and Leonardo Zannoli, former interns at the Experimental Lab (Department of Modern Languages Literatures and Cultures, University of Bologna), for their help in annotating the English, Italian and Spanish data.

References

- Amiot, D. and Van Goethem, K. 2012. A constructional account of French *-clé* ‘key’ and Dutch *sleutel-* ‘key’ as in *mot-clé* / *sleutelwoord* ‘key word’. *Morphology* 22: 347-364.
- Baayen, H. and Lieber, R. 1991. Productivity and English derivation: A corpus-based study. *Linguistics* 29: 801-843.
- Baayen, H. 2009. Corpus linguistics in morphology: Morphological productivity. In *Corpus linguistics. An international handbook*, A. Lüdeling and M. Kytö (eds), 900-919. Berlin and New York: de Gruyter.
- Bakema, P. and Geeraerts, D. 2000. Diminution and augmentation. In *Morphology. An International Handbook on Inflection and Word Formation*, G. Booij *et al.* (eds), 1045-1052. Berlin: de Gruyter.
- Barðdal, J. 2008. *Productivity: Evidence from case and argument structure in Icelandic*. Amsterdam and Philadelphia: John Benjamins.

- Bauer, L. 1983. *English word-formation*. Cambridge: Cambridge University Press.
- Bauer, L. 1997. Evaluative morphology: In search of universals. *Studies in Language* 21(3): 533–575.
- Blank, A. 2001. 112. Pathways of lexicalization. In *Language Typology and Language Universals*, Vol. II, M. Haspelmath, E. König, W. Oesterreicher and W. Raible (eds), 1596-1608. Berlin and Boston: De Gruyter.
- Booij, G. 2010. *Construction Morphology*. Oxford: Oxford University Press.
- Cappelle, B., Daus, R. and Hartmann, S. 2023. The English privative prefixes *near-*, *pseudo-* and *quasi-*: Approximation and ‘disproximation’. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1): 52–75.
- De Smet, H. 2012. The course of actualization. *Language* 88(3): 601–633.
- Dressler, W. U. and Barbaresi, L. M. 1994. *Morphopragmatics. Diminutives and intensifiers in Italian, German, and other languages*. Berlin and New York: de Gruyter.
- Efthymiou, A. 2015. Modern Greek diminutive and augmentative adjectives (in a cross-linguistic perspective). *SKASE Journal of Theoretical Linguistics* 21(1): 57–71.
- Eitelmann, M., Haugland, K. E. and Haumann, D. 2020. From *engl-isc* to *whatever-ish*: A corpus-based investigation of *-ish* derivation in the history of English. *English Language and Linguistics* 24(4): 801–831.
- Eitelmann, M. and Haumann, D. 2023. Getting *close-ish*: A corpus-based exploration of *-ish* as a marker of approximation and vagueness. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1): 76–100.
- Giannouloupoulou, G. 2006. On the borderline between lexicon and grammar: confixes in Modern Greek and Italian. *Language Typology and Universals* 59(3): 270–283.
- Gardani, F. 2020. Borrowing matter and pattern in morphology. An overview. *Morphology* 30: 263–282.

- Grandi, N. and Körtvélyessy, L. (eds). 2015. *Edinburgh Handbook of Evaluative Morphology*. Edinburgh: Edinburgh University Press.
- Grandi, N. 2017. I diminutivi come marche di attenuazione e indeterminatezza. In *L'expression de l'imprécision dans les langues romanes*, O.D. Balaş *et al.* (eds), 162–175. Bucharest: Ars docendi - Universitatea din Bucureşti.
- Haeringen, C. B. van 1956. *Nederlands tussen Duits en Engels*. The Hague: Servire.
- Hüning, M., Vogl, U., van der Wouden, T. and Verhagen, A. 2006. Nederlands tussen Duits en Engels. *Handelingen van de workshop aan de Freie Universität Berlin*. Leiden: Stichting Neerlandistiek Leiden.
- Iacobini, C. 2015. Foreign word-formation in Italian. In *Word-Formation: An International Handbook of the Languages of Europe*, Müller, P.O. *et al.* (eds), Vol. III, 1660–1679. Berlin: de Gruyter.
- Jurafsky, D. 1996. Universal tendencies in the semantics of the diminutive. *Language* 72(3): 533–578.
- Kastovsky, D. 2009. Astronaut, astrology, astrophysics: About combining forms, classical compounds and affixoids. In *Selected Proceedings of the 2008 Symposium on New Approaches in English Historical Lexis (HEL-LEX 2)*, McConchie, Roderick W. *et al.* (eds), 1–13. Somerville, MA: Cascadilla
- Kempf, L. and Eitelmann, M. 2018. Von diutisk zu dynamisch, von englisc zu anything-ish. is(c)h kontrastiv diachron. *Zeitschrift für Wortbildung / Journal of Word Formation* 2(1): 93–134.
- Kilgarriff, A. *et al.* 2014. The Sketch Engine: ten years on. *Lexicography* 1: 7–36.
- König, E. and Gast, V. 2018. *Understanding English-German Contrasts*. Fourth edition. Berlin: Erich Schmidt Verlag.
- Körtvélyessy, L. and Štekauer, P. (eds). 2011. Diminutives and augmentatives in the languages of the world. *Lexis: e-journal in English lexicology* 6: 5–25.

- Körtvélyessy, L. 2015. *Evaluative morphology from a cross-linguistic perspective*. Cambridge: Cambridge Scholars Publishing.
- Kuzmack, S. 2007. *Ish: A new case of antigrammaticalization*. Ms., University of Chicago.
- Lamiroy, B. 2011. Degré de grammaticalisation à travers les langues de la même famille. *Mémoires de la Société de linguistique de Paris* 19. 167–192.
- Levshina, N. 2015. *How to do Linguistics with R. Data exploration and statistical analysis*. Amsterdam and Philadelphia: John Benjamins.
- Masini, F. and Micheli, S. 2020. The morphological expression of approximation: the emerging *simil-* construction in Italian. *Word Structure* 13(3): 371–402.
- Masini, F., Norde, M. and Van Goethem, K. 2023. Approximation in morphology: A state of the art. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1): 1–26.
- Merlini Barbaresi, L. 2015. Evaluative morphology and pragmatics. In *Edinburgh Handbook of Evaluative Morphology*, N. Grandi and L. Körtvélyessy (eds), 32–42. Edinburgh: Edinburgh University Press.
- Norde, M. 2009. *Degrammaticalization*. Oxford: Oxford University Press.
- Norde, M. and Van Goethem, K. 2015. Emancipatie van affixen en affixoiden. *Nederlandse Taalkunde* 20. 109–148.
- Norde, M. and Van Goethem, K. 2018. Debonding and clipping of prefixoids in Germanic: Constructionalization or constructional Change? In *The Construction of Words*, G. Booij (ed.), 475–518. Cham etc.: Springer.
- Oltra-Massuet, I. 2017. Towards a morphosyntactic analysis of *-ish*. *Word Structure* 10(1). 54–78.
- Prčić, T. 2005. Prefixes vs initial combining forms in English: a lexicographic perspective. *International Journal of Lexicography* 18(3): 313–334.
- Prieto, V. M. 2005. *Spanish evaluative morphology: Pragmatic, sociolinguistic, and semantic issues*. Doctoral dissertation. Gainesville: University of Florida.

- Rainer, F. 2015. Intensification. In *Word-formation: An international handbook of the languages of Europe*, Müller, P. O. et al. (eds), vol. 2, 1339–1351. Berlin: de Gruyter.
- Schneider, K. P. 2003. *Diminutives in English*. Tübingen: Niemeyer.
- Seifart, F. 2015. Direct and indirect affix borrowing. *Language* 91(3): 511–532.
- Van Goethem, K. and De Smet, H. 2014. How nouns turn into adjectives. The emergence of new adjectives in French, English and Dutch through debonding processes. *Languages in Contrast* 14(2): 251–277.
- Van Goethem, K. and Koutsoukos, N. 2022. How typology shapes the constructional network: Denominal verb constructions in English, Dutch and German. *Zeitschrift für Wortbildung/Journal of Word Formation* 6(1): 7–57.
- Van Goethem, K. and Norde, M. 2020. Extravagant “fake” morphemes in Dutch. Morphological productivity, semantic profiles and categorical flexibility. *Corpus Linguistics and Linguistic Theory* 16(3): 425–458.
- Vassiliadou, H., Gerhard-Krait, F., Fotiadou, G. and Lammert, M. 2023. *Pseudo(-)* in French and Greek: Categorization and approximation. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1): 234–262.

Address for correspondence

Kristel Van Goethem

Institut Langage et Communication, Pôle de Linguistique, VALIBEL

F.R.S.-FNRS & Université catholique de Louvain

Place Blaise Pascal 1, L3.03.33

1348 Louvain-la-Neuve

Belgium

kristel.vangoethem@uclouvain.be

<https://orcid.org/0000-0003-2576-9011>

Co-authors information

Muriel Norde

Nordeuropa-Institut

Humboldt-Universität zu Berlin

muriel.norde@hu-berlin.de

<https://orcid.org/0000-0003-4944-8810>

Francesca Masini

Dipartimento di Lingue, Letterature e Culture Moderne (LILEC)

Alma Mater Studiorum – Università di Bologna

francesca.masini@unibo.it

<https://orcid.org/0000-0002-2354-4612>