ZWJW

Zeitschrift für Wortbildung
Journal of Word Formation

Sonderheft / Special Issue

Approximation in Morphology

Francesca Masini ● Muriel Norde ● Kristel Van Goethem (eds.)

herausgegeben von / edited by

ZW
JW

1/2023

HRSGG. / EDS.

Petra M. Vogel
Ludwig M. Eichinger
Mechthild Habermann
Jörg Meibauer
Barbara Schlücker
Hans-Joachim Solms
Pavol Štekauer
Salvador Valera Hernández

ZW
JW

Table of Contents

Francesca Masini, Muriel Norde & Kristel Van Goethem

Approximation in morphology:
A state of the art

Abstract: This paper offers a state of the art of approximation within the larger domain of evaluative morphology. It provides an overview of the formal means employed by the morphology of different languages to express approximative meanings, as well as a survey of the specific approximative values that can be conveyed. We further discuss the (input and output) lexical categories involved in the expression of approximation crosslinguistically. On an intralinguistic level, we address the issue of competition between different strategies.

Keywords: evaluative morphology, approximation, vagueness, categorization, competition, crosslinguistic variation

1. What is approximative morphology?[*]

1.1 Evaluative morphology and approximation

Evaluative morphology is by now a well-established domain of investigation (cf. e.g., Bauer 1997; Grandi & Körtvélyessy 2015; Körtvélyessy 2015). However, the semantic functions performed by evaluative morphemes have not been investigated in equal measure. Whereas functions such as diminution, augmentation and intensification have been studied quite extensively (cf., among many others, Dressler & Merlini Barbaresi 1994; Jurafsky 1996; Grandi 2002; Schneider 2003; Bakema & Geeraerts 2004; Prieto 2005; Körtvélyessy & Štekauer (eds.) 2011; Efthymiou 2015; Efthymiou, Fragaki & Markos 2015; Napoli 2017; Hendrikx 2019), others, like 'approximation', have received much less attention (see Amiot

---

& Stosic 2022 for a similar argument). Consequently, we still know very little about how approximation works within morphology.

1.2 Previous research

Only very recently has approximative morphology started to attract the attention of scholars. The most studied approximative morpheme is definitely English *-ish*, which has undergone remarkable semantic and morphological developments in the course of time. In present-day English, approximative *-ish* expresses vagueness and speaker attitude, and it has been expanding its host classes from adjectives (e.g., *warmish*) to a variety of other categories (e.g., *9-ish*, *okay-ish*) (e.g., Oltra-Massuet 2017; Kempf & Eitelmann 2018; Eitelmann, Haugland & Haumann 2020; Eitelmann & Haumann 2023). It has also been extensively discussed in the context of degrammaticalization because of its recent autonomous or "debonded" uses (Kuzmack 2007; Norde 2009), as in (1).

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

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 & Micheli 2020), and on 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 & Norde 2020). A large-scale cross-linguistic study is Van Goethem, Norde & Masini (2021), which is a corpus analysis of *pseudo-* in eight European languages (Danish, Dutch, English, German, French, Italian, Spanish and Swedish), with the aim to compare their morphological and distributional properties (e.g., Italian *pseudotifosi* 'pseudo-supporters'; Spanish *pseudoartistas* 'pseudo-artists'; German *pseudoreligiös* 'pseudo-religious'; Dutch *pseudo-wetenschappen* 'pseudo-sciences').

The fact that research into approximation is less established within morphological studies is also reflected by the abundance of terms that are used to refer to it, and by the lack of consensus on their use. For example, Grandi & Körtvélyessy (2015: 9, 11) use the triad 'approximation/reduction/attenuation' to refer to this functional domain. Other relevant terms are 'deintensification' (Körtvélyessy 2015: 63), 'non-prototypicality' (Cúneo 2015: 630), or 'non-authenticity', 'fakeness' and 'imitation' (Masini & Micheli 2020). Some of

these terms are typically associated with specific domains. For instance, 'attenuation' typically refers to 'reduced degree of a quality' in relation to adjectives (cf. e.g., Bauer 2002), like in French *blanchâtre* 'whitish'. Whereas 'attenuation' is rather widespread, terms like 'reduction', 'non-prototypicality' or 'deintensification' appear to be less common.

For this Special Issue, we decided to use **approximation** as a cover term for the (complex) functional domain associated with all these values: among the many available terms, it seems to be the one broadest in scope and, at the same time, one of the least associated with a specific domain/phenomenon. Contrary to Rainer (2015: 1346), who includes 'approximation/attenuation' in the wider category of 'intensification' (which "comprises not only a high degree but all degrees of intensity"), we keep approximation separate from intensification, even though the two domains are obviously closely related by being two opposite poles of the wider function of degree modification or reference modulation.

While approximation is relatively understudied in morphology, it has been at the center of interest of other subfields of linguistics, such as pragmatics and discourse studies, which produced a huge amount of literature on the topic. In these fields, other terms are more common, some of which identify specific types of approximation. The most used terms are 'mitigation' (Caffi 2007), 'indeterminacy' (Bazzanella 2011), 'imprecision' (Balaş et al. eds. 2017), 'hedging' (Lakoff 1972; Kaltenböck, Mihatsch & Schneider 2010), 'vagueness' (Channell 1994; Mihatsch 2007), 'intentional vagueness' (Voghera 2012; Voghera & Collu 2017), or 'defectiveness' (Amiot & Stosic 2022). These studies, however, have not been paying much attention to morphological means to convey approximation (compared to other strategies such as discourse markers and particles), with the exception of diminutive markers used as attenuation strategies or being derived from approximative values (cf. Merlini Barbaresi & Dressler 1994; Merlini Barbaresi 2015; Grandi 2017). Therefore, there seems to be an overlapping zone between studies on approximation in general and evaluative morphology (see Fig. 1) that is in need of further and deeper investigation.
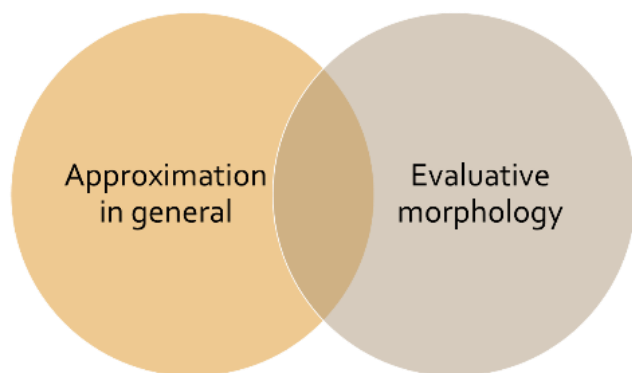
Fig. 1: Evaluative morphology and approximation studies

1.3 Research questions for this Special Issue

This Special Issue intends to bridge the gap between approaches to approximation in other subfields in linguistics and evaluative morphology, by concentrating on the expression of approximation by means of dedicated morphological means. The papers in this Special Issue are empirical studies of approximation in word formation from various theoretical perspectives, addressing one or more of the following research questions, each of which will be briefly discussed and illustrated in the following sections:

1. Which morphological means or **forms** are used to express approximating values?
2. What are the **sources** of approximating morphological markers crosslinguistically?
3. Which approximating **values** are expressed by morphological means?
4. Do we find **competition** between approximating morphological markers in a single language?
5. Do we find **crosslinguistic** tendencies/similarities in the morphological marking of approximation?

2. Forms

2.1 Morphological expression

As regards the morphological expression of approximation, **suffixation** (2) and **prefixation** (3) seem to be the predominant processes in the case studies in this Special Issue.

(2) a. Sicilian *longa* 'long' > *lungareḍḍa* 'longish' (Brucale & Mocciaro 2023)

b. Italian *attimo* 'instant' > *attimino* 'instant.DIM' (Voghera 2023)

c. English *warm* > *warmish* (Eitelmann & Haumann 2023)

d. French *jaune* 'yellow' > *jaunâtre* 'yellowish' (Stosic & Amiot 2023)

e. Dutch *migraine* 'migraine' > *migraineachtig* 'migraine-like' (Hüning & Schlücker 2023)

f. Kambaata   *maɾf-**lab**-á*   *ut-ichch-ú*

             needle-APRX-M.ACC   thorn-SGV-M.ACC

             'a thorn-like needle'   (Treis 2023)

(3) a. Italian *besciamella* 'béchamel' > ***para**-besciamella* 'béchamel-like' (Micheli 2023)

b. Greek *επιστήμη* 'science' > *ψευτο-επιστήμη* (***psefto**epistími*) 'pseudo-science' (Vassiliadou et al. 2023)

c. French *classique* 'classical' > ***pseudo**-classique* 'pseudo-classical' (Vassiliadou et al. 2023)

d. English *diplomatic* > ***quasi**-diplomatic* (Cappelle, Daugs & Hartmann 2023)

More marginally, **circumfixation** may also be used to express approximation, as can be seen in example (4) from Georgian.

(4)   Georgian *mžave* 'sour' > ***mo**-mžav-**o*** 'slightly sour' (Topadze Gäumann 2015: 221–222)

In addition to affixes, **compound stems** or **affixoids** may also be used to express approximative values, as shown in the examples under (5).

(5) a. Italian *simil-* 'similar': ***simil**-coppia* 'pseudo-couple' (Masini & Micheli 2020)

b. French *-forme* 'form': *une substance géli**forme*** 'a gel-like substance' (Stosic & Amiot 2023)

c. Dutch *imitatie* 'imitation': ***imitatie**leer* 'imitation leather' (Van Goethem & Norde 2020)

Finally, **reduplication** can be a morphological means to express approximation, as illustrated by examples (6a–b) (quoted in Masini & Di Donato 2023). Apart from contiguous reduplication, we find **discontinuous reduplication**, as in (6c) (Masini & Di Donato 2023).

(6) a. Makasar *leʼleng* 'black' > *leʼleng-leʼleng* 'blackish' (Jukes 2006: 105)

b. Maʼdi *īŋgwē* 'white' > *īŋgwē īŋgwē* 'whiteish [sic], rather white' (Blackings & Fabb 2003: 105)

      c.   Italian   *sapone-non-sapone*   lit.   soap-NEG-soap   'soap-free   cleanser' (Masini & Di Donato 2023)

2.2 Parts of speech as input and output categories

With respect to the bases (lexical categories) that serve as **input** of the morphological processes mentioned in the preceding section, the examples given thus far already show that nouns and adjectives are the most common bases. More examples are given in (7) for **nouns** and (8) for **adjectives**. However, approximative morphology can also be found in **verbs** (9), albeit to a lesser extent.

(7)   a.   Italian *attore* 'actor' > *attore-non-attore* actor-NEG-actor 'unconventional actor' (Masini & Di Donato 2023)

      b.   French *frère* 'brother' > *frérot* 'bro' (Stosic & Amiot 2023)

(8)   a.   English *quasi-diplomatic* (Cappelle, Daugs & Hartmann 2023)

      b.   Italian *semi-aperto* 'partially open' (Micheli 2023)

(9)   a.   Greek *ψευτοκοιμάμαι* (pseftokimáme) 'I pseudo-sleep' (Vassiliadou et al. 2023)

      b.   Italian *quasi-cercare* 'sort of look for' (Masini & Micheli 2020)

More marginally still, approximative morphemes may apply to **adverbs** (10a) and **prepositions** (10b):

(10)   a.   Sicilian *cchiossai* 'more' > *cchjessanieḍḍu* 'a little more' (Brucale & Mocciaro 2023)

      b.   Italian *simil-contro* 'sort of against' (Masini & Micheli 2020)

It is worth noting that many strategies allow more than one base, like for instance English *near*, which attaches to both nouns (*near-perfection*) and adjectives (*near-fatal*; Cappelle et al. 2023). English *-ish* (Eitelmann & Haumann 2023) is probably the most flexible approximative morpheme in this respect. In addition to all previous base types, it may attach to full phrases (11a) or even completely debond (11b) (cf., among others, Norde 2009; Van Goethem & Norde 2020). This debonded use is however not exclusive for *-ish*. As shown in (11c), the neoclassical combining form *pseudo-*, for instance, also undergoes debonding in different languages, such as Swedish.

(11)   a.   English *four years ago-**ish*** (Eitelmann & Haumann 2023)

      b.   English *I can cook. **Ish**.* (Eitelmann & Haumann 2023)

    c.   Swedish *[…] urbaniteten i fråga […] blev först **pseudo**, sen sjangserade, och sen efterhärmades* 'The urbanity under discussion […] first became pseudo, then faded and then it was being imitated' (Van Goethem, Norde & Masini 2021)

As regards the **output** categories of the morphological processes, homocategoriality between input and output category seems to be prevalent (in line with evaluative morphology properties; Grandi & Körtvelyessy 2015), especially with approximating prefixes as left constituents (12).

(12)   a.   French *enquête* 'investigation' > *pseudo-enquête* 'pseudo-investigation' (N > N) (Vassiliadou et al. 2023)

        b.   Italian *scientifico* 'scientific' > *parascientifico* 'parascientific' (A > A); *scheggiarsi* 'chip' > *semi-scheggiarsi* 'nearly chip' (V > V) (Micheli 2023)

According to Stosic & Amiot (2023), homocategoriality is even defined as a constraint for diminutives to express approximation, see the opposition between (13a) and (13b):

(13)   a.   French *livre* 'book' > *livret* 'small book' (N > N)

        b.   French *bleu* 'blue' > *bleuet* 'cornflower' (A > N)

However, word-class change in approximative word-formation is far from excluded, as shown in the contributions by Hüning & Schlücker (2023) on denominal adjectives in Germanic (14a) and by Treis (2023) on the derivational suffix *-lab* in Kambaata, which converts nouns, verbs and adjectives into adjectives (14b).

(14)   a.   English *peacock* > *peacock-**like** feathers* (N > A)

        b.   Kambaata *torr-* 'throw' > *torr-**lab**-á* 'in a kind of throwing way' (V > A)

The English *-ish* morpheme again stands out here since it can take scope over a noun phrase (15) (Eitelmann & Haumann 2023).

(15)   GEIST: *How long have you been together?*
           KOTB: [[*A year and a half*]$_{DP}$ -**ish**]$_{DP}$. [COCA]

3. Sources

The approximating strategies illustrated in the previous section happen to derive from different sources that will be discussed in this section. What follows is a representative –

although not necessarily exhaustive – list of possible sources, most of which are explored in the articles hosted in this Special Issue, whereas others still need further exploration.

– **Negation items**: negative prefixes like *non-* in French (and possibly other languages) may convey approximation (e.g., *non-ville* 'non-city', namely a city with non-stereotypical properties, cf. Dugas 2017), but negative markers may also have a role in larger structures, like the Italian N-*non*-N discontinuous reduplication analysed by Masini & Di Donato (2023) in this volume (e.g., *sapone-non-sapone*, lit. soap-NEG-soap, 'soap-free cleanser').

– **Fake items**: items that convey meanings such as fakeness, imitation, simulation/pretending may easily turn into approximators. A representative example is the already mentioned *pseudo-*, from Ancient Greek *pseudēs* 'false', which is further analysed in this volume by Cappelle, Daugs & Hartmann (2023) and Vassiliadou et al. (2023). A possible example for the 'pretending' type is Finnish *-vinA* (*maksa-vina-an*, lit. pay-VINA-POSS.3 '(they) pretend to pay') (Salminen 2000).

– **Degree and quantity items**: this class can be split into at least two subclasses, namely: (i) items conveying an 'almost' or incompleteness meaning, e.g., *quasi-* in various European languages (e.g., English *quasiparticle*) (Cappelle, Daugs & Hartmann 2023); (ii) items meaning 'half', like Italian *semi-* (*semi-relazione* 'pseudo-relationship'; Micheli 2023) or *mezzo-* (*mezzo-pacifiste* 'half-pacifists'; Masini & Micheli 2020; Benigni 2022).

– **Spatial (proximity) items**: items that convey closeness may also become approximators, see for instance English *near-* (*near-synonyms*, *near-identical*; cf. Cappelle, Daugs & Hartmann 2023) or the neoclassical form *para-* (from Greek *para* 'beside'), studied for Italian by Micheli (2023).

– **Diminutives**: the relation between diminutives and approximation is well-known (Jurafsky 1996); indeed, diminutives (unlike augmentatives, cf. Stosic & Amiot 2023) are commonly used to attenuate the meaning of the base (Dressler & Merlini Barbaresi 1994; Merlini Barbaresi 2015; Grandi 2017; also Voghera 2023), with different nuances depending on the lexical category and semantics of the base, cf. English *-let* in *star* (N) > *starlet* ' a young actor likely to develop into a star' (Dixon 2014: 172); French *-ette* in *réforme* 'reform' (N) > *réformette* 'small reform with no value or scope' (Amiot & Stosic

2022); Sicilian *-ḍḍa* in *longa* 'long' (ADJ) > *lungareḍḍa* 'a bit long' (Brucale & Mocciaro 2023); Italian *-erellare* in *cantare* 'sing' (V) > *canterellare* 'to sing a bit/carelessly' or *-cchiare* in *bruciare* 'burn' (V) > *bruciacchiare* 'burn slightly' (Grandi 2008; cf. Audring, Leufkens & van Lier 2021 for a typological overview of verbal diminutives).

– **Similative items**: elements meaning 'like/as', 'seem/resemble' or 'similar/alike' are commonly used as sources for approximation in the world's languages (cf. Masini, Micheli & Huang 2018 for an overview); this is unsurprising given that "the semantic transition from similative comparison markers to approximation is extremely easy" (Mihatsch 2009: 79) since "similative comparison is never perfect, but always approximative" (Mihatsch 2009: 70, quoting Haspelmath & Buchholz 1998: 278). German *-ähnlich* '-like' (*ähnlich* is an adjective meaning 'similar'; Hüning & Schlücker 2023) and English *-like* (*baptismal-like, prefix-like*) are good examples. Bauer, Lieber & Plag (2013: 311–313) mention *-like* among similative adjective-deriving affixes (together with *-esque*, *-ish* and *-oid*) that have an approximating reading derived by inference from the similative one. Other examples come from Kambaata (Cushitic), e.g., *-lab*, from a verb root *lab-* 'resemble, seem' (e.g., *gamball-(i)-lab-á(ta)* 'blackish') (Treis 2023), and Georgian (Kartvelian), e.g., *-savit*, which includes the postposition *vit* 'like/as' (e.g., *davighale-savit* PRV-get_tired.AOR.1SUB-ES 'I got like tired'; Topadze Gäumann 2015: 223).

– **Relational items**: similarly to similative items, relational items – which express association and therefore, to some extent, similarity and belonging – may turn into approximators as well: this is the history Eitelmann & Haumann (2023) reconstruct for English *-ish*, which develops its current approximative meaning (*cleanish*, *50-ish*) from a more relational/associative one (*Spanish*, *heavenish*).

– **Taxonomic items**: these are possibly the best-known sources for approximation, giving rise to well-known markers of the *sort*/*kind of* type in a variety of languages (e.g., among many others, Tabor 1994; Denison 2002; Mihatsch 2007; Traugott 2008; Voghera 2013, 2017; Masini 2016). Taxonomic items also have a role in the evolution of morphological affixes, as is the case for German *-artig* (e.g., *vogelartiges Tier* 'bird-like animal'), whose literal meaning is 'belonging to a species ('Art')' (Hüning & Schlücker 2023).

– **Modal items**: finally, it is worth mentioning a source domain which is still underinvestigated, namely modality. In English and Italian, for instance, some epistemic adverbs are developing an approximative meaning in compound-like expressions: see, e.g., English *maybe* in *a maybe-incident* or Italian *forse* 'maybe' in *forse-fidanzato* 'maybe-boyfriend'. The same is true of volitional expressions, like English *wannabe* (from *want to be*) in *wannabe popstar*, German *möchtegern* 'would very much like to' in *Möchtegern-Schriftsteller* 'wannabe author' or Italian *vorrei-ma-non-posso* '(I) would like but (I) can't' in *boutique vorrei-ma-non-posso* 'wannabe boutique' (cf. Norde et al. 2023). The conceptual step here is from an assessment of speakers' (un)certainty about what they say (epistemic modality) or an aspiration to attain a specific property/state (volitional modality) to approximation.

4. Values

As may already be evident from the many examples given in the preceding sections, approximation is not a clear-cut domain and we would not expect anything different. As Ullmann (1962: 118) writes about the closely related notion of 'vagueness': "If one looks more closely at this vagueness one soon discovers that the term is itself rather vague and ambiguous". The same can be said of approximation, which appears to be a complex functional domain comprising closely related yet distinct values, spanning from semantics to pragmatics.

In this section we identify and discuss some of these values (without any claim to exhaustiveness), focusing on those that emerge from the papers in the Special Issue, with special attention to approximating values that affect the propositional content (namely, what Voghera & Borges 2017 call "informational vagueness" *vs.* "relational/discourse vagueness"). As we will see, what emerges is that specific values may sometimes be quite difficult to define and that many markers may convey more than one value, which may generate quite some confusion.

The first notion that we consider is **privativity**,[1] which Cappelle, Daugs & Hartmann (2023) define as a function that removes an essential property from an entity X, with the result that the output Y does not belong to X's category anymore ('not X'). For instance, *fake* in *fake blood* would have a privative function since *fake blood* does not match the properties of real blood, and therefore it falls out of the blood category. The privative function can therefore be regarded as the most extreme value of approximation, which implies the creation of a new (X-related) category rather than the modulation or evaluation of an existing category. See also the diachronic approximation cline proposed by Eitelmann & Haumann (2023), running from relational to privative (which is the meaning English *-ish* does not seem to realise).

Another value to be considered is **disproximation**, which is elaborated by Cappelle, Daugs & Hartmann (2023) in order to better characterize the (different) semantic contribution carried by the prefixes *near-*, *pseudo-* and *quasi-* in English. The authors advocate for a 'dynamic' or 'orientational' approach to meaning, which is implemented through Langackerian representations: whereas *quasi-* and especially *near-* are 'approximating', namely coming close to a standard (coinciding with the base X), *pseudo-* is '*dis*proximating' in that it falls short of something with respect to the standard, conveying an idea of deficiency which is not prominent in *quasi-*/*near-*.

In their semantic analysis of *pseudo-* in French and Greek, Vassiliadou et al. (2023) acknowledge its privative/disproximating value (category exclusion), but they also define two other non-privative values that *pseudo-* may convey, thus producing a three-way classification: the first coincides with a **subjective depreciation**, which implies category inclusion (see French *pseudo-supporteur* 'pseudo-supporter', denoting a team supporter who is a bad exemplar of the category for some reason and therefore judged negatively by the speaker); the second is vagueness intended as **undecidable categorization**, where category membership cannot be ascertained (see French *pseudo-liberté* 'pseudo-freedom'). Category inclusion or exclusion is indeed a recurrent criterion throughout the papers in the Special Issue. See also Brucale and Mocciaro's (2023) proposal to distinguish between:

---

[1] Note that, in the typological literature, the term 'privative' normally refers to lexical and grammatical markers of absence like without or -less (cf. Stolz, Stroh & Urdze 2007), another equivalent term being 'caritive' (Oskolskaya et al. 2020).

(i) **internal approximation**, which entails a semantic modification of the base without altering category membership (see Sicilian *chilu* 'kilo' > *chiliceḍḍu* 'about one kilo'; *muzzucuni* 'bite' > *muzzucunieḍḍu* 'small bite'); (ii) **external approximation**, where the deviation from the standard is such that category membership is affected (see Sicilian *acqua* 'water' > *acquiceḍḍa* 'sort of water', but also more lexicalized cases like *figura* 'figure' > *figureḍḍa* 'holy card').

Next, **attenuation** (or **reduction**) is probably one of the most widespread and quoted functions in the approximation domain. It is intended as the reduced degree of a quality or property and is especially common with color terms (French *verdâtre* 'greenish', English *whitish*, Italian *marroncino* 'brownish', etc.) and other gradable adjectives (see (16a, c)). However, this function also applies to nouns (16b). In addition to reduced degree, attenuation may have to do with partialness or incompleteness, as the translations of the Italian examples (16a–b) illustrate.

(16) a. Italian *addormentato* 'asleep' > **semi**-*addormentato* 'partially asleep' (Micheli 2023)

  b. Italian *oscurità* 'darkness' > **semi**-*oscurità* 'partial darkness' (Micheli 2023)

  c. Kambaata *mux-i-**láb**-a-a<n>ta*
      wet-EP-**APRX**-F.PRED-F.COP2<EMP>
      'a bit wet, almost dry'   (Treis 2023)

Other values that are worth discussing are **fakeness**, **imitation** and **simulation**. As mentioned in Section 3, fake items have a role in the development of approximators, but **non-authenticity** can also be regarded as a specific value of approximation. The already mentioned *fake blood* is not a type of blood but still resembles blood because it reproduces some of its properties. On the other hand, as also pointed out by Cappelle, Denis & Keller (2018) (to which we refer for a finer-grained typology of 'fakeness'), a *fake article* is still a type of article, despite deviating from the normal article in that its content is not true (as it should be). The question arises whether we are dealing with one single value (say, 'non-authenticity') or a family of related values. In their study of Italian *simil-*, Masini and Micheli (2020: 383) distinguish two different values: (i) fakeness, which is realized when "Y refers to an entity that is meant to imitate X (without being a genuine X) in order to be taken as an X" (e.g., *similpelle* 'imitation leather'); (ii) imitation, which is realized when "Y refers to an entity that is meant to imitate or reproduce X (without being a genuine X), the

difference with [fakeness] being that there is no intent for Y to be taken as an X" (e.g., *simil-vita*, lit. SIMIL-life, referring to the virtual life of a videogame character). So, the distinction lies in the criterion of intentionality, which may produce different effects in terms of connotation. Both fakeness and imitation, as intended by Masini and Micheli (2020), are closely related to what we may call the **simulative** (or 'pretending') function, which is exemplified by the Finnish *-vina-* construction discussed by Ebner (2022) and exemplified in (17):[2]

(17) | *Pitä-ä* | *osa-ta* | *teeskennel-lä* | *että* | *ol-isi* | *jotenkin* |
|---|---|---|---|---|---|
| must-3SG | be_able-INF | pretend-INF | that | be-COND | somehow |

*"teke-**vinä**-än*	*työ-tä."*
do-VINA-POSS.3SG	work-PART

'One has to be able to pretend as if one was "seemingly working" somehow [although in reality there is no work]'

Beside fakeness and imitation, Masini & Micheli (2020) identify three other functions expressed by Italian *simil-*, namely: resemblance, (intentional) vagueness and kin-categorization.

Kin-categorization is realized when Y results in a separate category (with some sort of link to X) that gets conventionalized to a certain extent (e.g., Italian *simil-matrimonio*, lit. SIMIL-marriage 'civil union'). This value has clear connections to both privativity (Cappelle, Daugs & Hartmann 2023) and external approximation (Brucale and Mocciaro 2023).

Resemblance, or **similarity**, is a central value that is instantiated by a variety of means and can be regarded as more basic and neutral than other values, because – in a way – approximation always involes *some* (level/type of) similarity between X and Y. It is not surprising, then, that resemblance or similarity is called upon, as a value, in many contributions within the Special Issue. If we take resemblance to be the value that emerges when "Y refers to an entity that is merely similar to X without being an X" (Masini & Micheli 2020: 383; e.g., Italian *simil-Amazon* 'SIMIL-Amazon', referring to an account that resembles the Amazon account but is not Amazon, namely an 'Amazon-style' account),

---

[2] Ebner's (2022) study of Finnish offers further insight into the (extension of the notion) of approximation. Beside the *-vina* construction, Ebner (2022) mentions the *-maisilla-* construction, which may define still another dimension of approximation, namely approximation as a function of time. Indeed, the *-maisilla-* construction is normally described in terms of "proximative aspect" (Ylikoski 2003: 43). The relation between (imminent or proximative) aspect and approximation definitely deserves further thought and investigation.

then we are in the domain of privativity and/or external approximation, although – as we already noted – category membership is not always so easy to determine. In her analysis of Kambaata's marker -*lab*, Treis (2023) further distinguishes between "similarity and near-identity in manner" (e.g., *uucc-* 'beg' > *uucc-**lab**-á* 'in a kind of begging way') and "similarity in quality or character" (e.g., *weteechchu* '*weteechchu* tree' > *wet-eechch-**lab**-á* '*weteechchu*-like'). Sometimes, resemblance is very close to imitation and distinguishing between the two values can be quite difficult (cf. Masini & Micheli 2020 on Italian *simil-* and Micheli 2023 on Italian *semi-*).

Hüning & Schlücker (2023) add yet another value to the picture, namely **comparison** (which has obviously to do with similarity). An -*achtig* formation in Dutch like *leerachtig* 'leather-like' in *een leerachtige substantie* 'a leathery/leather-like substance' can be both approximative (the substance has leather-like properties but is not leather) and comparative (the substance has leather-like properties regardless of its relation to the category leather), and differentiating between the two is not always easy or possible. The authors therefore recommend to distinguish approximative markers in the strict sense (which are privative and typically left-headed) from comparative markers with an optional approximative reading.

Last but not least, one of the main values expressed is (**intentional**) **vagueness** (Voghera 2012, 2013; Voghera & Collu 2017). This term refers to vagueness intentionally expressed by the speaker, not vagueness that depends on systemic factors (like *bald* being systemically a "vague predicate" because it gives rise to borderline cases, unlike for instance *prime*, when referring to numbers). Intentional vagueness has to do with uncertain or fuzzy categorization, with reference to entities that are (possibly) peripheral members of a category and whose categorial status is unclear. We already discussed the value of **undecidable categorization** as conceptualized by Vassiliadou et al. (2023). This may indeed be one subtype of intentional vagueness, the other being **non-prototypicality**, which expresses the peripheral status of an entity with respect to prototypical exemplars, due to some deviating property or properties to be reconstructed by the interlocutor (cf. Masini & Di Donato 2023). Non-prototypicality may result in category inclusion or exclusion, depending on individual cases.

Interestingly, expressions conveying intentional vagueness developed also a focus function, as illustrated for Italian diminutives by Voghera (2023). Thus, in example (18), the vague diminutive *un attimino* 'an instant.DIM' puts the following NP ('this draft law that is being discussed') in focus.

(18)   *vi volevo aggiornare **un attimino** su questo disegno di legge che è in discussione*
          'I wanted to update you a little instant on this draft law that is being discussed'

In the background of all these values lies a bigger question: What is the **relationship between approximation and categorization**? Where does approximation end and where does categorization start (or vice versa)? When we use expressions like *blueish* or *prefix-like* are we approximating an existing category or concept or rather creating a new one? In response to a question like *What's your favourite shade of blue?*, we are likely to reply something like *baby blue* or *dark blue*, but not *blueish*. Still, *blueish* has a conceptual value, closely related to *blue*, although possibly peripheral with respect to the prototypical core of the category. Whereas the concept *baby blue*, as a specific shade of *blue*, is probably more stable and shared by a large number of people, *blueish* is more variable and subject to speakers' interpretation. Moreover, it probably blurs into other closely related concepts like for instance *greenish*, due to the fuzzy boundaries of color categories (see Fig. 2).
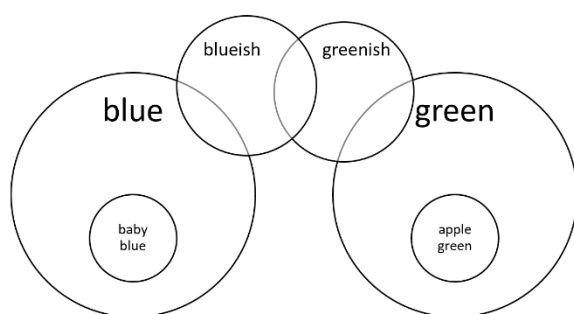
Fig. 2: Blueish vs. baby blue

To sum up, we may try to visualize the – admittedly complex – picture described in this section as in Fig. 3. The figure schematizes the *continuum* going from category modulation or 'evaluation' (here, specifically, through approximation) to (new) category creation.
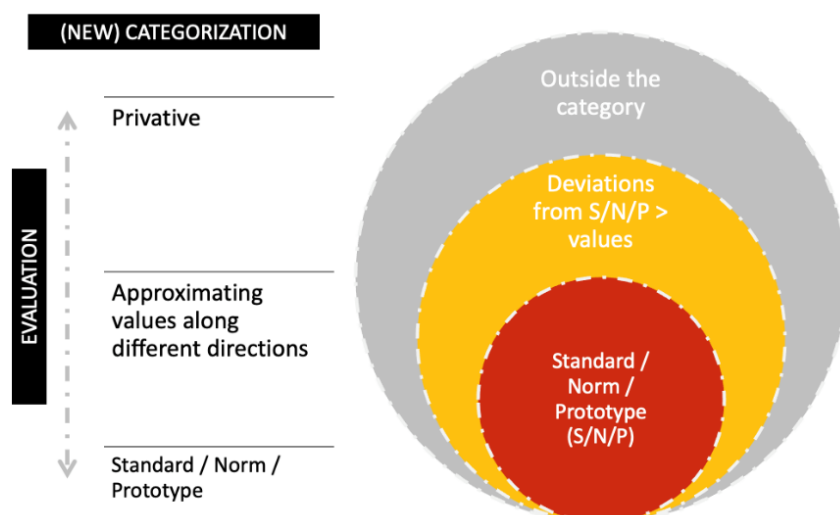
Fig. 3: From approximation to categorization

The boundaries between the three areas – red (indicating the standard/norm/prototype, which coincides with the base X), yellow (representing the deviations from the standard/norm/prototype) and grey (indicating the conceptual area beyond the category X) – are intended as fuzzy (hence the dotted white line), in line with prototype semantics: it is not always clear if an approximated item Y still belongs to the standard's category X or not. This representation is in harmony with: (i) the view of evaluation (and hence also approximation) as a process that expresses a deviation from a standard or default value (Grandi 2017); (ii) the view that regards approximation and categorization as two sides of the same coin (cf. Vassiliadou et al. 2023; Vassiliadou & Lammert 2022). Finally, deviations from the standard (the yellow area) is purposely oversimplified in Fig. 3: as is clear from the above discussion, deviations may be of different types and intensity, generating a range of approximating values, also depending on the nature of the base X.[3] Hence, more fine-grained schemas are needed for each value, some of which are provided by the papers in this Special Issue. Note that the standard/norm/prototype may be recursive: *baby blue*, for instance, is a hyponym of *blue*, which may well develop its own deviations (e.g., *babyblueish*, *near-babyblue*, etc.).

---

[3] One direction that is not explored in the Special Issue but would be worth pursuing is the hypothesis that some sources (Section 3) may tendentially develop into approximation markers of some sort only, thus constraining the limits of possible diachronic development. According to Masini, Micheli & Huang (2018), approximation markers derived from similative items tend to convey only a subset of the possible values: they can end up expressing intentional vagueness or imitation, whereas attenuation seems to be excluded or dispreferred (a claim to be tested on a wider sample of languages).

5. Competition

Approximative constructions often find themselves in competition with each other, at different levels.

First of all, we may find different structural strategies competing with each other, for instance X-*wannabe* (*actress wannabe*) versus *wannabe*-X (*wannabe actress*) (Norde et al. 2023).

Secondly, different approximative morphemes can be found with the same base, as is shown by examples (19a) and (19b), in which the German approximative suffixes *-ähnlich* and *-artig* both collocate with *Honig* 'honey' to form denominal adjectives (Hüning & Schlücker 2023).

(19)   a.   *Eine **honigähnliche** Zuckerpaste wird gegen die natürliche Wuchsrichtung des Haares aufgetragen und dringt so bis zum Haarschaft ein.*

'A honey-like sugar paste is applied against the natural direction of hair growth, penetrating all the way to the hair shaft.'

b.   *Wer es süß mag, kann die frischen Blüten zu einem Gelee oder einem **honigartigen** Sirup für einen Brotaufstrich verarbeiten.*

'For those who like it sweet, the fresh flowers can be made into a jelly or a honey-like syrup for a spread.'

In other cases, competing morphemes occur in very similar collocations, for instance in the food items in (20) (Eitelmann & Haumann 2023).

(20)   *Is everything in here onion-**like**, bacon-**esque** or cheese-**ish**?*

Nevertheless, competing approximative morphemes, even when they appear in similar contexts, may show distributional preferences (Hüning & Schlücker 2023) and/or semantic differences. For example, Bauer, Lieber & Plag (2013: 416) argue that the prefixes *pseudo-* and *quasi-* and the suffixes *-ish*, *-esque* and *-oid* all denote "something that is similar, but not identical, in shape or quality to what the base denotes", although only *pseudo-* conveys a notion of 'falseness'. As far as the suffixes are concerned, they claim that these may differ according to register or domain without further specification. But as Cappelle, Haugs & Hartmann (2023) show, it is possible to reveal differences using methods from distributional semantics, such as semantic vector spaces (for another semantic vector

analysis, of Dutch compounding elements meaning 'fake', see Van Goethem & Norde 2020).

In the paper by Treis in this Special Issue, on the other hand, the (near-)synonymy of approximative morphemes is being disputed. Treis (2023) quotes a Kambaata dictionary entry (Alemu 2016: 340) containing a compounding element *manka'-á* 'humility, behavior, manner(s), norm(s)', whose definition contains the approximative suffix *-lab*, as illustrated in (21). However, most of the dictionary entries with *manka'-á* were unknown to Treis' language assistant, nor did they appear in elicited data, so either there is no competition, or there is but is limited to specific varities of Kambaata.

(21) *baar-**manka'**-á*   [Definition:] *feeg-á*   *baar-á*
      yellow-**manner**-M.ACC   bright-M.ACC   yellow-M.ACC

      ih-umb-ú;   *baar-**lab**-á*
      become-3M.NEG5-M.ACC   yellow-**APRX**-M.ACC

      'yellowish' [Definition:] 'not being bright yellow, yellowish'

A third level of competition is one between different variants of the same morpheme, such as Modern Greek *pseudo-* and *psefto-* (Vassiliadou et al. 2023) in *ψευτο/ψευδο-επιστήμη* (psefto/psevðo-epistími) 'pseudo-science' or *ψευτο/ψευδοδίλημμα* (psefto/psevðoðílima) 'pseudodilemma'. As Vassiliadou et al. (2023) show in their analysis of Modern Greek corpus data, *pseudo-* and *psefto-* do however show distributional differences, e.g., in the parts of speech they collocate with (with more variation for *psefto-*).

Finally, approximative morphemes may compete with other (non-morphological) constructions, e.g., expressions such as English *kind of/kinda* and *sort of/sorta* (Eitelmann & Haumann 2023), or the Sicilian degree adverb *tanticchia/nna picca* 'a bit' (Brucale & Mocciaro 2023).

Factors determining these various kinds of competition are largely a topic for future research, but early studies indicate that strongly lexicalized approximative items may block the use of other approximators, as in the case of Dutch *kunstmest* 'fake manure > fertilizer', which blocks the use of words like *ˀimitatiemest* ('imitation manure'). What is most probably at play here is what Goldberg (2016: 377–378) has termed "statistical pre-emption", which she defines as "a particular type of indirect negative evidence that results from repeatedly hearing a formulation, B, in a context where one might have expected to

hear a semantically and pragmatically related alternative formulation, A." The choice between competing approximative morphemes can also be constrained by semantic factors. Thus, the sense of *pseudo-* of 'pretending to be without actually being' precludes collocations such as ?*pseudo-fatal* (Cappelle, Haugs & Hartmann 2023). Yet other factors remain to be explored. These include the etymology of the base (e.g., neo-classical compounds vs native bases) or register.

6. Cross-linguistic observations

The studies in this Special Issue are drawn from a fairly large (albeit unbalanced) sample of languages: Germanic (English, Dutch, German), Romance (Italian, French, Sicilian), Greek, Serbian – all IE languages – and Kambaata (Cuschitic). From this sample, some very general cross-linguistic tendencies can be derived:

i)   morphological means to express approximation seem to be widely spread across European languages and beyond;

ii) there is a tendency for approximative affixes to accommodate different kinds of bases, especially nouns, adjectives and verbs, but other categories are not excluded;

iii) the types of sources are also quite recurrent among languages, with single languages often displaying multiple markers from different sources;

iv) languages often display an array of approximating values, sometimes conveyed by the same marker(s), sometimes expressed through dedicated strategies; disentangling the exact relationship between forms and functions and how they impact the choice between competing expressions seems to be one of the most challenging tasks that lie ahead.

The similarity displayed by European languages in the morphological expression of approximation is also due to their common heritage from classical languages (see neoclassical combining forms such as *pseudo-*, *quasi-*, *para-*, *semi-*, etc.). Nevertheless, as shown in the cross-linguistic study by Van Goethem, Norde & Masini (2021) on the prefix *pseudo-*, the same approximative morpheme may have different fate and behavior cross-linguistically. Although that study indicates that *pseudo-* displays a similar lexical distribution in Romance and Germanic languages, notable differences in productivity, for

instance, are detected. Fig. 4 shows the "global productivity" (Baayen and Lieber 1991: 818–819) of *pseudo-* in the eight languages under examination, with potential productivity (i.e. the number of hapaxes divided by the number of tokens) 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 red-dotted, the latter blue-dotted). It can be derived from the graph that the differences in productivity are not only related to the language group (overall higher productivity in Romance than in Germanic) but also to the base types (overall higher productivity in nominal than in adjectival formations).
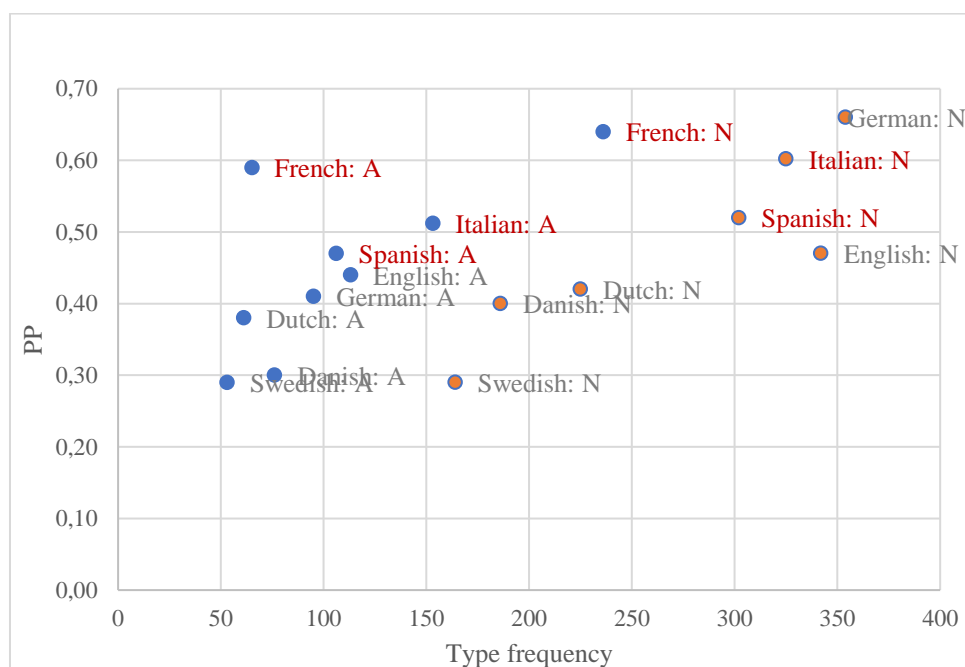
Fig. 4: Global productivity of *pseudo-*

To explain these kinds of differences, intralinguistic competition is one of the factors to be taken into account (cf. also Cappelle, Daugs & Hartmann 2023). Crucially, there is not always a one-to-one mapping between approximative morphemes in different languages. The contribution by Hüning & Schlücker (2023) nicely illustrates this point for Germanic denominal suffixation: it shows that three different German approximative suffixes correspond to Dutch *-achtig* (and English *-like*), namely *-artig*, *-ähnlich* and *-haft*. These compete with each other and show a similar high frequency of use, albeit with important lexical-distributional differences. More fine-grained contrastive analyses of this kind are

needed to further explore the universal and divergent tendencies in the field of approximative morphology from a cross-linguistic perspective.

References

Alemu Banta Atara [Alamu Banta Ataara]. 2016. *Kookaata: Kambaatissa-Amaarsa-Ingiliizissa laaga doonnuta* [Kambaata-Amharic-English dictionary]. Addis Ababa: Berhanena Selam Printing.

Amiot, Dany & Dejan Stosic. 2022. Evaluative morphology: From evaluation to approximation and semi-categorization. In Héléne Vassiliadou & Marie Lammert (eds.), *A Crosslinguistic Perspective on Clear and Approximate Categorization*, 53–94. Newcastle upon Tyne: Cambridge Scholars Publishing.

Audring, Jenny, Sterre Leufkens & Eva van Lier. 2021. Small events: Verbal diminutives in the languages of the world. *Linguistic Typology at the Crossroads* 1(1). 223–256.

Baayen, Harald & Rochelle Lieber. 1991. Productivity and English derivation: A corpus-based study. *Linguistics* 29. 801–843.

Bakema, Peter & Dirk Geeraerts. 2000. Diminution and augmentation. In Geert Booij, Christian Lehmann, & Joachim Mugdan (eds.), *Morphology. An International Handbook on Inflection and Word Formation*, 1045–1052. Berlin & New York: de Gruyter.

Bauer, Laurie. 1997. Evaluative morphology: In search of universals. *Studies in Language* 21(3). 533–575.

Bauer, Laurie. 2002. What you can do with derivational morphology. In Sabrina Bendjaballah, Wolfgang U. Dressler, Oskar E. Pfeiffer & Maria D. Voeikova (eds.), *Morphology 2000: Selected papers from the 9th Morphology Meeting, Vienna, 24–28 February 2000*, 37–48. Amsterdam & Philadelphia: John Benjamins.

Bauer, Laurie, Rochelle Lieber & Ingo Plag. 2013. *The Oxford Reference Guide to English Morphology.* Oxford: Oxford University Press.

Balaş, Oana Dana, Adriana Ciama, Mihai Enăchescu, Anamaria Gebăilă & Roxana Voicu (eds.). 2017. *L'expression de l'imprécision dans les langues romanes*. Bucharest: Ars docendi - Universitatea din București.

Bazzanella, Carla. 2011. Indeterminacy in dialogue. *Language and Dialogue* 1. 21–43.

Benigni, Valentina. 2022. *È stato un mezzo disastro…* Fenomeni di convergenza tra vaghezza e intensificazione: un approccio contrastivo tra italiano e russo. Paper presented at "Italiano e lingue slave: confronto, contatto, insegnamento", Ss. Cyril and Methodius University in Skopje, 23–24 June 2022.

Blackings, Mairi & Nigel Fabb. 2003. *A Grammar of Ma'di*. Berlin: Mouton de Gruyter.

Brucale, Luisa & Egle Mocciaro. 2023. Approximation through suffixation: *-ḍḍu/-a* in Sicilian. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1). 27–51.

Caffi, Claudia. 2007. *Mitigation*. Amsterdam: Elsevier.

Cappelle, Bert, Robert Daugs & Stefan Hartmann. 2023. The English privative prefixes *near-*, *pseudo-* and *quasi-*: Approximation and 'disproximation'. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1). 52–75.

Cappelle, Bert, Pascal Denis & Mikaela Keller. 2018. Facing the facts of fake: A distributional semantics and corpus annotation approach. Yearbook of the German Cognitive Linguistics Association 6(1). 9–42.

Channell, Joanna M. 1994. *Vague Language*. Oxford: Oxford University Press.

Cúneo, Paola. 2015. Toba. In Nicola Grandi & Lívia Körtvélyessy (eds.), *The Edinburgh Handbook of Evaluative Morphology*, 625–633. Edinburgh: Edinburgh University Press.

Denison, David. 2002. History of the *sort of* construction family. Paper presented at the Second International Conference on Construction Grammar (ICCG2), Helsinki, September 6–8, 2002.

Dixon, R. M. W. 2014. *Making New Words: Morphological Derivation in English*. Oxford: Oxford University Press.

Dressler, Wolfgang U. & Lavinia M. Barbaresi. 1994. *Morphopragmatics. Diminutives and Intensifiers in Italian, German, and Other Languages*. Berlin & New York: Mouton de Gruyter.

Dugas, Edwige. 2017. Constraints on the French [*non*-N] construction. In Nikos Koutsoukos, Jenny Audring & Francesca Masini (eds.), *Morphological Variation: Synchrony and Diachrony. Online Proceedings of the Eleventh Mediterranean Morphology Meeting (MMM11)*, Nicosia, Cyprus, 22–25 June 2017, 14–25.

Ebner, Daniel. 2022. Approximation in Finnish verbal morphology. Paper presented at "ApproxiMo: Approximation in Morphology" (online), 8 February 2022.

Efthymiou, Angeliki. 2015. Modern Greek diminutive and augmentative adjectives (in a cross-linguistic perspective). *SKASE Journal of Theoretical Linguistics* 21(1). 57–71.

Efthymiou, Angeliki, Georgia Fragaki & Angelos Markos. 2015. Exploring the polysemy of the Modern Greek prefix *iper-*. *Morphology* 25(4). 411–438.

Eitelmann, Matthias, Kari. E. Haugland & Dagmar Haumann. 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, Matthias & Dagmar Haumann. 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). 80–104.

Goldberg, Adele E. 2016. Partial productivity of linguistic constructions: Dynamic categorization and statistical preemption. *Language and Cognition* 8. 369–390.

Grandi, Nicola. 2002. *Morfologie in contatto: le costruzioni valutative nelle lingue del Mediterraneo*. Milan: FrancoAngeli.

Grandi, Nicola. 2008. *I verbi deverbali suffissati in italiano*. Roma: Caissa Italia.

Grandi, Nicola. 2017. I diminutivi come marche di attenuazione e indeterminatezza. In Oana Balaş, Adriana Ciama, Mihai Enăchescu, Anamaria Gebăilă, Roxana Voicu (eds.), *L'expression de l'imprécision dans les langues romanes*, 162–175. Bucharest: Ars docendi - Universitatea din Bucureşti.

Grandi, Nicola & Lívia Körtvélyessy. 2015. Introduction: Why evaluative morphology? In Nicola Grandi & Lívia Körtvélyessy (eds.), *The Edinburgh Handbook of Evaluative Morphology*, 3–20. Edinburgh: Edinburgh University Press.

Haspelmath, Martin & Oda Buchholz. 1998. Equative and similative constructions in the languages of Europe. In Johan van der Auwera (ed.), *Adverbial Constructions in the Languages of Europe*, 277–334. Berlin & New York: De Gruyter.

Hendrikx, Isa. 2019. *The acquisition of intensifying constructions in Dutch and English by French-speaking CLIL and non-CLIL students: Cross-linguistic influence and exposure effects*. PhD dissertation. Louvain-la-Neuve: Université catholique de Louvain.

Hüning, Matthias & Barbara Schlücker. 2023. Approximation and comparison in word-formation. The case of denominal adjectives in Dutch, German, and English. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1). 101–129.

Jukes, Anthony. 2006. *Makassarese (basa Mangkasara'): A Description of an Austronesian Language of South Sulawesi*. University of Melbourne PhD dissertation.

Jurafsky, Dan. 1996. Universal tendencies in the semantics of the diminutive. *Language* 72(3). 533–578.

Kaltenböck, Gunther, Wiltrud Mihatsch & Stefan Schneider. 2010. *Introduction*. In Gunther Kaltenböck, Wiltrud Mihatsch & Stefan Schneider (eds.), *New Approaches to Hedging*, 1–13. Bingley: Emerald.

Kempf, Luise & Matthias Eitelmann. 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.

Körtvélyessy, Lívia. 2015. *Evaluative Morphology from a Cross-linguistic Perspective*. Cambridge: Cambridge Scholars Publishing.

Körtvélyessy, Lívia & Pavol Štekauer (eds.). 2011. Diminutives and augmentatives in the languages of the world. *Lexis: e-journal in English lexicology* 6. 5–25.

Kuzmack, Stefanie. 2007. *Ish:* A new case of antigrammaticalization. Paper presented at the Annual Meeting of the Linguistic Society of America, Anaheim, CA, 4–7 January 2007.

Lakoff, George. 1973 [1972]. Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic* 2. 458–508.

Masini, Francesca. 2016. Binominal constructions in Italian of the N1-di-N2 type: Towards a typology of Light Noun Constructions. *Language Sciences* 53. 99–103.

Masini, Francesca & Jacopo Di Donato. 2023. Non-prototypicality by (discontinuous) reduplication: The N-*non*-N construction in Italian. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1). 130–155.

Masini, Francesca & M. Silvia Micheli. 2020. The morphological expression of approximation: The emerging *simil-* construction in Italian. *Word Structure* 13(3). 371–402.

Masini, Francesca, M. Silvia Micheli & Shanshan Huang. 2018. Attenuation and categorization: A cross-linguistic study. Paper presented at the Workshop "Attenuated qualities in a cross-linguistic perspective", SLE 51, Tallinn (Estonia), 29 August – 1 September 2018.

Merlini Barbaresi, Lavinia. 2015. Evaluative morphology and pragmatics. In Nicola Grandi & Lívia Körtvélyessy (eds.), *The Edinburgh Handbook of Evaluative Morphology*, 32–42. Edinburgh: Edinburgh University Press.

Micheli, M. Silvia. 2023. The emergence of approximative values in the Italian prefixes *semi-* and *para-*. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1). 156–176.

Mihatsch, Wiltrud. 2007. The construction of vagueness: "sort of" expressions in Romance languages. In Günther Radden, Klaus-Michael Köpcke, Thomas Berg & Peter Siemund (eds.), *Aspects of Meaning Construction*, 225–245. Amsterdam: Benjamins.

Mihatsch, Wiltrud. 2009. The approximators French *comme*, Italian *come*, Portuguese *como*, Spanish *como* from a grammaticalization perspective. In Corinne Rossari, Claudi Ricci & Adriana Spiridon (eds.), *Grammaticalization and Pragmatics*, 65–92. Bingley: Emerald.

Napoli, Maria. 2017. Nomi in *stra-* in italiano: intensificazione tra semantica e pragmatica. In Alain Lemaréchal, Peter Koch & Pierre Swiggers (eds.), *Actes du XXVIIe Congrès international de linguistique et de philologie romanes*, 95–105. Nancy: ATILF.

Norde, Muriel. 2009. *Degrammaticalization*. Oxford: Oxford University Press.

Norde, Muriel, Francesca Masini, Kristel Van Goethem & Daniel Ebner. 2023. *Wannabe* approximatives: creativity, routinization or both? Paper presented at the Workshop "AG 1: Dynamics at the lexicon-syntax interface: Creativity and routine in word-formation and multi-word expressions", 45th Annual Meeting of the German Linguistic Society (DGfS2023), Cologne, 8–10 March 2023.

Oltra-Massuet, Isabel. 2017. Towards a morphosyntactic analysis of *-ish*. *Word Structure* 10(1). 54–78.

Oskolskaya, Sofia A., Natalia M. Zaika, Sergei B. Klimenko & Maksim L. Fedotov. 2020. Определение каритива как сравнительного понятия [Defining caritive as a comparative concept]. *Voprosy Yazykoznaniya* 3. 7–25.

Prieto, Victor Moises. 2005. *Spanish Evaluative Morphology: Pragmatic, Sociolinguistic, and Semantic Issues*. PhD dissertation. Gainesville: University of Florida.

Rainer, Franz. 2015. Intensification. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-formation: An International Handbook of the Languages of Europe, vol. 2*, 1339–1351. Berlin & New York: De Gruyter.

Salminen, Taru. 2000. *Morfologiasta moniäänisyyteen. Suomen kielen kvasirakenteen merkitys, käyttö ja kehitys*. Helsinki: Suomalaisen Kirjallisuuden Seura.

Schneider, Klaus. P. 2003. *Diminutives in English*. Tübingen: Niemeyer.

Stolz, Thomas, Cornelia Stroh & Aina Urdze. 2007. WITH(OUT): On the markedness relation between comitatives/instrumentals and abessives. *Word* 58. 63–122.

Stosic, Dejan & Dany Amiot. 2023. Can suffixal augmentation express approximation? The case of evaluative nouns. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1). 177–203.

Tabor, Whitney. 1994. The gradual development of degree modifier *sort of* and *kind of*. In Katharine Beals, Gina Cooke, David Kathman, Sotaro Kita, Karl-Erik McCullough & David Testen (eds.), *Proceedings of the 29th Regional Meeting of the Chicago Linguistic Society*, 451–465. Chicago: Chicago Linguistic Society.

Topadze Gäumann, Manana. 2015. Georgian. In Nicola Grandi & Lívia Körtvélyessy (eds.), *The Edinburgh Handbook of Evaluative Morphology*, 216–226. Edinburgh: Edinburgh University Press.

Traugott, Elizabeth C. 2008. The grammaticalization of NP of NP constructions. In Alexander Bergs and Gabriele Diewald (eds.), *Constructions and Language Change*. Berlin & New York: de Gruyter.

Treis, Yvonne. 2023. The approximative derivation in Kambaata (Cushitic). *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1). 204–233.

Ullmann, Stephen. 1962. *Semantics: An Introduction to the Science of Meaning*. Oxford: Blackwell.

Van Goethem, Kristel & Muriel Norde. 2020. Extravagant "fake" morphemes in Dutch. Morphological productivity, semantic profiles and categorical flexibility. *Corpus Linguistics and Linguistic Theory* 16(3). 425–458.

Van Goethem, Kristel, Muriel Norde & Francesca Masini. 2021. The fate of 'pseudo-' words: A contrastive corpus-based analysis. Paper presented at UCCTS 2021 (Using Corpora in Contrastive and Translation Studies), Bertinoro, 9–11 September 2021.

Vassiliadou, Hélène, Francine Gerhard-Krait, Georgia Fotiadou & Marie Lammert. 2023. *Pseudo*(-) in French and Greek: Categorization and approximation. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1). 234–262.

Vassiliadou, Hélène & Marie Lammert. 2022. Clear and approximate categorization: An overview. In Hélène Vassiliadou & Marie Lammert (eds.), *A Crosslinguistic Perspective on Clear and Approximate Categorization*, 1–22. Newcastle upon Tyne: Cambridge Scholars Publishing.

Voghera, Miriam. 2012. Chitarre, violini, banjo e cose del genere. In Anna M. Thornton & Miriam Voghera (eds.), *Per Tullio De Mauro: studi offerti dalle allieve in occasione del suo 80° compleanno*, 341–364. Roma: Aracne.

Voghera, Miriam. 2013. A case study on the relationship between grammatical change and synchronic variation: The emergence of tipo$_{[-N]}$ in Italian. In Anna Giacalone Ramat, Caterina Mauri & Piera Molinelli (eds.), *Synchrony and Diachrony. A Dynamic Interface*, 283–312. Amsterdam & Philadelphia: Benjamins.

Voghera, Miriam. 2017. La nascita delle costruzioni non nominali di *specie*, *genere*, *sorta* e *tipo*: uno studio basato su corpora. In Paolo D'Achille & Maria Grossmann (eds.), *Per la storia della formazione delle parole in italiano*, 277–307. Firenze: Franco Cesati.

Voghera, Miriam. 2023. The role of evaluative suffixes in vagueness and focus construction: The Time Nouns constructions. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1). 263–286.

Voghera, Miriam & Carla Borges. 2017. Vagueness expressions in Italian, Spanish and English task-oriented dialogues. *Normas* 7(1). 57–74.

Voghera, Miriam & Laura Collu. 2017. Intentional vagueness: A corpus-based analysis of Italian and German. In Maria Napoli & Miriam Ravetto (eds.), *Exploring Intensification: Synchronic, Diachronic and Cross-linguistic Perspectives*, 371–388. Amsterdam & Philadelphia: Benjamins.

Ylikoski, Jussi. 2003. Havaintoja suomen ns. viidennen infinitiivin käytöstä. *Sananjalka* 45. 7–44.

Francesca Masini
Alma Mater Studiorum – Università di Bologna
Dipartimento di Lingue, Letterature e Culture Moderne
Via Cartoleria 5
40124 Bologna, Italy
francesca.masini@unibo.it

Muriel Norde
Nordeuropa-Institut
Humboldt-Universität zu Berlin
Unter den Linden 8
10099 Berlin, Germany
muriel.norde@hu-berlin.de

Kristel Van Goethem
F.R.S.-FNRS & Université catholique de Louvain
Institut Langage et Communication / PLIN / VALIBEL
Place Blaise Pascal 1
1348 Louvain-la-Neuve, Belgium
kristel.vangoethem@uclouvain.be

Luisa Brucale & Egle Mocciaro

Approximation through suffixation:
-ḍḍu/-a in Sicilian

Abstract: This article investigates the use of the Sicilian suffix *-ḍḍu/-a* for the expression of approximation. On the basis of a survey of a corpus of ethnotexts and the outputs of a translation questionnaire, we propose that approximation is a core value in the semantic network of the suffix, expressing a certain distance from the default values conveyed by the base. In terms of prototypicality, this distance may occur from the categorial centre (internal approximation): the suffix modifies the semantics of the base but does not alter the categorial status of the referent. Alternatively, the suffix may impact on categorial membership tout court (external approximation), questioning the categorial properties of the base. Based on this classification, the semantic cores of the suffix are represented in a map showing the connections between the various meanings identified in the corpus.

Keywords: Sicilian evaluative suffixes, approximation, attenuation, high specificity, diminution, pragmatic nuances

1. Introduction[*]

This article explores a little investigated phenomenon in the existing descriptions of Sicilian, namely the use of evaluative suffixes for the expression of approximation (cf. classical accounts such as De Gregorio 1907; Rohlfs 1969: 374–375, 402–404; Varvaro 1988: 724). Only recently, Emmi (2011: 267, our translation) has pointed out "an attenuative meaning [...] tending towards approximation" for some evaluative suffixes. A case in point is *-ḍḍu/-ḍḍa* (-/ḍḍu/, /-ḍḍa/), as in (1) and (2):

(1)   *Tinia a criata pi ffàricci coccu **ssuvviżżieḍḍu*** [ALS 36, 199]

      '(S)he kept a maid for her to do a few **minor housework**'

(2)   *Nun jera cafè, era **acquiceḍḍa** marruncina* [from the questionnaires]

'It was not coffee; it was **a sort of** brownish **water**'

In (1) and (2), rather than expressing smallness, the suffix appears to attenuate the categorial membership properties of the output noun (respectively, a 'not really demanding housework' and 'a sort of water, difficult to recognise as coffee').

The use of -*ḍḍu*/*ḍḍa* is investigated by analysing a corpus consisting of a selection of ethnotexts published by the *Centro di studi filologici e linguistici siciliani* and the answers to a translation-type questionnaire administered to speakers from different parts of Sicily. Data analysis is conducted against a framework which combines the polysemic model developed within cognitive linguistics (in particular, Jurafsky 1996) and the morpho-pragmatic approach to diminutives (Dressler & Merlini Barbaresi 1994; Grandi 2002, 2011, 2015, 2017 inter alia; Grandi & Körtvélyessy 2015). In this perspective, diminutive suffixes express a subjective evaluation that is not necessarily diminutive in a concrete sense. Based on this, we propose that the central value in the semantic network of the suffix -*ḍḍu* is approximation. In general terms, this notion refers to the state of proximity to a reference value, where exact knowledge of the degree of distance between an approximating entity and an approximated entity is lacking (see, for example, Hamawand 2017). When expressed through morphological means, approximation can be understood more precisely as a certain degree of distance from the default values expressed by the base. This definition can be rearticulated in terms of prototypicality and, on this basis, we propose that -*ḍḍu* expresses two levels of approximation:

a) Distance from the categorial centre (*internal approximation* or *attenuation*). The suffix modifies the semantics of the base X, but does not alter the categorial status of the referent: X-*ḍḍu* is a non-prototypical example of X, but it's still a X. A case in point is *ssuvviżżieḍḍu* 'minor housework' in (1); other examples are *chiliceḍḍu* 'about one kilo (a little less, a little more)' (< *chilu*) and *muzzucunieḍḍu* 'small bite' (< *muzzucuni*). All these cases are in fact hyponyms of X, whose categorical properties are *attenuated*.

b) Impact on categorial membership tout court (*external approximation*). The suffix questions the categorial properties of the base and this can occur at two levels. First, X-*ḍḍu* exhibits such specific characteristics as ends up being considered as an entity autonomous from X, e.g., *figureḍḍa* 'holy card' < *figura* 'figure', which conventionalises a

specific value other than X (even if still referable to it). Second, X-*ḍḍu* is linked to X only by a similarity relationship ('a sort of X, a X-like', but not a X). We have already seen *acquiceḍḍa* in (2); another example (with a greater degree of semantic autonomy since the derivative's meaning is conventionalised rather than context-dependent), is *lattucheḍḍa* 'seaweed' < *lattuca* 'lettuce'; in both cases the derivative forms can hardly be interpreted as hyponyms of the base.

In both a) and b), the categorial interpretation depends on the subjective perspective of the speakers; when subjectivity is no longer at work, we are dealing with lexicalised and opaque examples of the suffix, to be discussed separately.

The paper is organised as follows: in Sections 2 and 3, we report, respectively, on the etymology of the suffix and the existing studies on its uses in Sicilian; the theoretical premises on which the analysis is carried out are made explicit in Section 4, while Section 5 is devoted to the description of the corpus; the data analysis is carried out in the next Sections and includes a schematic representation of the possible types of approximation (Section 6) and the description of specific cases (Sections 6.1 to 6.3); a semantic map of the various meanings expressed by *-ḍḍu/ḍḍa* is proposed in Section 7, together with some final remarks.

2. Etymology

The Latin ancestor of *-ḍḍu/-a* is the suffix *-llus*, *-a*, *-um* (Lat. *-ll-* > Sic. *-ḍḍ-*), generally traced back to the Indo-European suffixes **lo* and **ko* (cf. Haverling 2011 inter alia). While the former was more frequent and multifunctional than the latter, they could combine in the form *-culus/-a/-um* in words of the third, fourth and fifth declension (e.g., *navicula* 'little boat' < *navis* 'boat'). In other cases, **lo* occurred alone, producing themes in *-olo-/ -ulo-* and in *-ll-*. Therefore, at least originally, *-culus*, *-olus*, *-ulus* and *-llus* were in complementary distribution, since the occurrence of one or the other was determined by paradigmatic facts, related to the inflectional class to which they belonged, and by syntagmatic facts, motivated by the phonotactic context. More precisely, *-llus* is formed by adding *-lo-* to lateral, vibrant or nasal roots; the contact between *-r*, *-n*, *-l* radicals and *-l* of the suffix results in an assimilation that produces the gemination of the lateral *-ll-*, while outcomes in *-e*, *-i* or in *-u* (i.e., *-ellus*, *-illus* or *-ullus*) depend on the root vowel, as in *libēllus* 'small

wrap, booklet' < *liber* 'book', *catīllus* 'bowl' < *catinus* 'plate, *satūllus* 'sated' < *satur* 'sated'. This distribution changes throughout the history of Latin, as *-llus* increases its productivity regardless of inflectional class (*novellus* 'new, young' < *novus* 'new', *nepotilla* 'niece' < *nepos* 'nephew' etc.).

On the semantic level, the Indo-European suffixes **lo* and **ko* and their Latin outcomes have been the subject of numerous hypotheses since Brugmann (1906), who argued that they originally had no diminutive meaning, but served to express similarity and/or belonging to a species or type. Brugmann's hypothesis is generally accepted in the relevant literature. For instance, Hakamies (1951) provided a list of words showing the use of the suffixes to form nouns and adjectives denoting relationships of similarity and belonging. Based on Brugmann (1906), Magni (1999) proposed that the two suffixes were markers of (inalienable) possession; this original semantic nucleus would then have been articulated in historical languages, including Latin, through metaphorical processes that produced relational marks of inclusion ('of the same species/type as X') and metonymic connections resulting in the expression of meronymic relations ('part/all'). Through these processes, the suffixes **lo* and **ko* lost their original motivation, spreading to wider domains that also include diminutive and hypocoristic uses. A different etymology for **ko*, which overturns Brugmann's approach, has been proposed by Jurafsky (1996), who, based on cross-linguistic observation, traced the suffix back to an original concrete meaning 'child' (not attested in historical Indo-European languages), from which all others would have irradiated out through abstraction processes of various kind; this model will be discussed, with a critical eye, in Section 4.

3. Existing studies on Sicilian diminutive suffixes

The first attempt to describe forms and uses of Sicilian diminutives is a work by De Gregorio (1907: 216), who listed the following suffixes: *-areḍḍu, -a*; *-eḍḍu, -a*; *-icchiu, -a*; *-iceḍḍu, -a*; *-iḍḍu, -a*; *-iteḍḍu, -a*; *-ittu, -a*; *-olu, -a*; *-uḍḍu, -a*; *-uliḍḍu, -a*; *-ulu, -a*; *-uzzu, -a*. According to De Gregorio, *-eḍḍu/-a* is the most used suffix to form Sicilian diminutives, even more frequent if we also include what he called "secondary formations", that is, ante-

suffixal interfixed derivatives such as *-ar-eḍḍu*, *-ic-eḍḍu*, *-it-eḍḍu*, *-ul-iḍḍu*, mainly involving bisyllabic words[1].

De Gregorio's description is supported by a more recent study by Emmi (2011: 240–260; 272–273). Based on wide sampling of modern and contemporary texts[2], she observed both the greater frequency of *-eḍḍu* compared to other suffixes and the preference of interfixes for bisyllabic bases (*panz-ar-eḍḍa*N 'little belly', *picc-ar-eḍḍa*Adv 'a little bit'). Furthermore, consistent with De Gregorio, she treats *-eḍḍu*, *-iḍḍu* and *-uḍḍu* separately rather than as allomorphs of the same morpheme (see also Rohlfs 1969: 402–404; Varvaro 1988: 724). In fact, sometimes the two forms appear as free variants added to the same base (cf. *mascul-eḍḍu*/*mascul-iḍḍu* 'little male, little boy'). However, this more frequently results in different words, as in *caruseḍḍu* 'moneybox' and *carusiḍḍu* 'little boy', which would suggest – Emmi claims – that *-eḍḍu* and *-iḍḍu* are different suffixes. Emmi (2011: 291) also notes the different behaviour of *-eḍḍu*/*-iḍḍu* with interfixes, which may combine with *-eḍḍu* and not with *-iḍḍu*, sometimes producing synonymous pairs such as *muss-iḍḍu*/*mussi-c-eḍḍu*, both of which lemmatised in Piccitto & Tropea's *Vocabolario Siciliano* (1985: III, 923) as diminutives of *mussu* 'snout, lips'.

Mainly based on the etymological reasons discussed in Section 2, in what follows, we will consider *-eḍḍu*, *-iḍḍu*, *-uḍḍu* as variants of the same suffix *-ḍḍu*. However, the distribution of these variants of the suffix undoubtedly still needs to be investigated.

4. Theoretical premises

The analysis proposed here is characterised by a radically semantic interpretation of the suffix and its uses, which also involves a polysemic approach to them (Grandi & Körtvélyessy 2015; Jurafsky 1996; Mutz 2015; Prieto 2015 inter alia). The starting point is the radial model proposed by Jurafsky (1996, based on Lakoff 1987) to describe the semantic network of diminutives, which is said to be organised around a centre, represented by

---

[1]As Merlini Barbaresi (2004: 276) points out, these phonic segments "are devoid of denotative meaning, even though they may contribute to autonomous connotative meanings and strategic uses in discourse".

[2] The corpus studied by Emmi (2011) contains the works in Sicilian by Luigi Pirandello (1915–1918), Luigi Capuana (1895) and Nino Martoglio (1903), the sectorial dictionaries of the *Atlante Linguistico della Sicilia* (ALS) and the *Vocabolario siciliano* by Piccitto and Tropea (1985).

the value 'child', and its most immediate extension, that is, 'small', from which other more peripheral values progressively radiate (see also Mutz 2015: 149). This network is shown in Fig. 1:
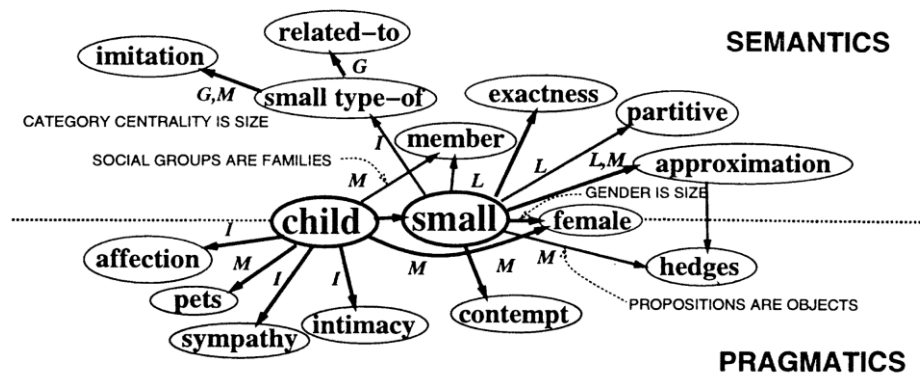
Fig. 1: Semantic network of diminutives (Jurafsky 1996: 542)

According to Jurafsky (1996: 537, 552, 565), Fig. 1 has crosslinguistic validity and shows both synchronic and diachronic relationships. In other words, where not attested, as in the case of Indo-European **ko* and its various historical developments, the diachronic priority of 'child' must still be reconstructed or inferred (Jurafsky 1996: 568–569). All other more abstract meanings would be later developments based on mechanisms such as metaphor (M), generalisation (G), conventionalisation of inferences (I) and what Jurafsky (1996: 554) called lambda-abstraction (L), an abstraction mechanism that quantifies not on objects but on predicates (e.g., *red → reddish*, where the suffixed adjective approximates the meaning of the unmarked base). However, as already noted by Magni (1999), data do not always support the diachronic priority of the 'child' meaning postulated by Jurafsky, nor do they explain the semantics of Indo-European and Latin diminutive suffixes. As for the latter, on the basis of the analysis of the suffix *-llus* in Plautus' comedies, Brucale & Mocciaro (2020; see also Prieto 2015: 26) argued for the synchronic centrality of the meaning 'small', which is actually a core value also in Jurafsky's map and one of the most commonly attested cross-linguistically. However, despite being well attested (55 out of 67 occurrences), this meaning hardly corresponds to a mere and concrete (i.e., physical) smallness in Plautus' corpus. Rather, *-llus* is always associated to a range of other 'small'-related senses, which Brucale & Mocciaro (2020: 78) arranged in a synchronic network incorporating both semantic and pragmatic values. Semantic values instantiate in two core areas, namely attenuation ('small or non-prototypical exemplar of') and approximation ('similar to, sort of'). The pragmatic

values (e.g., affective or playful attitude, emotion, familiarity, courtesy, euphemism, irony, modifications of the illocutionary force etc.; see Dressler & Merlini Barbaresi 1994: 173–380 and the subsequent morpho-pragmatic literature) are facets of the value 'small' in its attenuative specification or, to put it in another way, the attenuative value is never expressed in Plautus unless linked to a pragmatic dimension. What connects all this senses is a schematic meaning of 'generic relation', which acts in synchrony as the network's *fil rouge* (but could also have – as Brucale and Mocciaro 2020: 80 suggest – a diachronic validity in the spirit of Brugmann's proposal). In fact, the notion of 'small' is inherently relational since it implies the existence of related non-small entities (the unmarked or central examples of the category modified by the suffix). In other words, *-llus* may indicate 'a less prototypical exemplar of, a specimen subordinated to, a small version of; a sort of etc.' In this sense, 'small' is only one of the possible specifications of a meaning of 'generic relation', to which all attested meanings can be traced[3].

Brucale and Mocciaro's (2020) analysis, which is a premise for the study presented here, is supported by several studies on evaluative morphology which have assigned a central role to values such as attenuation, vagueness and imprecision in the description of the semantics of diminutives (e.g., Grandi & Körtvélyessy 2015; Grandi 2017). We will refer here mainly to Grandi (2017), who proposed two substantial revisions to Jurafsky's model. First, whereas Jurafsky (1996: 554) linked attenuative values to gradable concepts and thus primarily to adjectives and verbs, Grandi (2017: 146–147) observes that even nouns can be assigned an evaluative suffix that does not alter the referential ability of the base (e.g., It. *-ino* in *maritino* 'hubby' and *-accio* in *ragazzaccio* 'bad boy', which are still a 'husband' and a 'boy'). As Grandi (2017: 147–148, our translation) claims,

> any evaluative form with a compositional and transparent semantic reading, i.e., not lexicalised, is subordinate with respect to the base noun. In other words, any evaluative form designates an object, an action, a property, etc. with a certain degree of approximation with respect to what is expressed by the base noun, i.e., it represents an object, an action, a property, etc. in a way that is, so to speak, inaccurate.

---

[3] The schematic sense of 'generic relation' could also explain the 12 *-llus* words in the corpus that cannot be attributed to 'small', such as *capillus* 'hair' < *caput* 'head', *suillus* 'pertaining to, belonging to pigs' < *sus* 'pig', *flabellum* 'fan' < *flabrum* 'blow', all examples of ancient, sometimes morphologically opaque words, such as *stella* 'star' from the same root of *sternere* 'to spread, scatter' (Brucale & Mocciaro 2020: 61–62).

An internal articulation of the evaluative notions used emerges from Grandi's discourse. While every evaluation implies an *approximation*, the latter can be either by excess and thus encoded by augmentatives, or by defect and encoded by diminutives. This approximation by defect coincides, according to Grandi (2017: 148, our translation and emphasis), "with an *attenuation* of some characteristic attributed to the value identified as a norm or standard". In this sense, attenuation would be a specification of the broader category of approximation. Based on this premise, Grandi reverses the directionality of the 'small → attenuative' relationship postulated by Jurafsky (and this is the second fundamental revision to the radial model) and instead assigns a primary (archetypal) role to the latter (see also Hakamies 1951). This is explained as follows:

> [A] *tavolino* obviously indicates a small table; but a small table is nothing more than an approximation by default to the standard image of a table (within a specific culture, of course). Hence, in fact, an attenuation of the size property of a 'standard' table. [...] The decrease or, rather, the semantic function 'small' would in this view be a special case of attenuation, which, in turn, would be a special case of approximation. (Grandi 2017: 148, our translation)

Grandi's description also has an impact on the diachronic primacy of the value of 'child' postulated by Jurafsky. Although it is undoubtedly true that diminutives originating from forms indicating the parental relationship ('child, puppy of X') are attested in many Indo-European and other languages, the sense 'child' cannot necessarily be considered as originating. This is evident from the history of the Latin suffix *-īnus < *-īno-* (Grandi 2003: 122), which is the basis of the main Romance diminutives (e.g., It. *-ino*, Sp. *-ín*, Port. *-inho*) and which was used with a multiplicity of functions traced back to a generic relational value (similarly to what has been described for *-illus* above). Hence, a form such as *castŏrīnus*$_A$ would have had an original relational sense ('resembling a beaver'), then, via conversion, would have developed the meaning of 'young exemplar of a beaver', in line with Hakamies' (1951: 9) observation that "l'adulte est le prototype d'une espèce; par conséquent ce qui ou celui qui ressemble à l'espèce sans atteindre toutefois au prototype ne peut être que plus petit". In short, Grandi (2011: 16) reconstructed the following development: *castor*$_N$ 'beaver' > *castŏrīnus*$_A$ 'resembling a beaver > *castŏrīnus*$_N$ 'young beaver > *castorino*$_N$ 'small/little beaver', in which the value 'child' is said to occupy only an intermediate position in the development chain.

By recognising the primacy of approximation, Grandi overcomes the articulation into descriptive (i.e., "physical, hence therefore tangible, concrete, observable, testable" (2017: 141); in other words, *objective*) and qualitative (i.e., *subjective*) evaluation proposed in previous works (e.g., Grandi 2011 and Grandi & Körtvélyessy 2015: 10), on which also Emmi's (2011) classification of the Sicilian suffix is partly based), and traces back all possible 'deviations' from the default meaning of the unmarked base that are encoded by the evaluative suffix to the expression of the speaker's *subjectivity*. As Voghera (2023) explains, the subjective meanings expressed by diminutives can be arranged along a continuum that goes from a shared *intersubjectivity* to a *high subjectivity*:

Intersubjectively shared meaning ⟶	**Highly subjective meaning**
fazzolettino	*maritino*
'handkerchief+dim, tissue, kleenex'	'husband+dim, hubby'
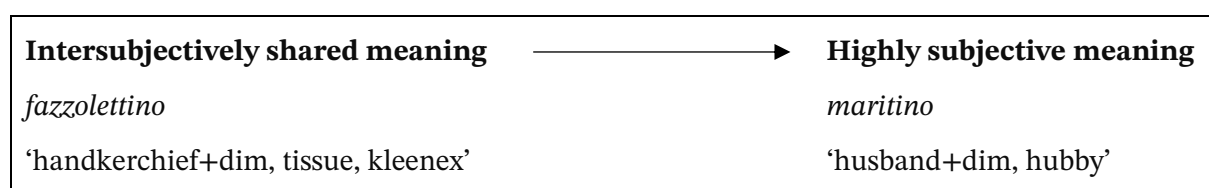
She clarifies that

> [i]f Italian speakers use the word *fazzolettino* ('handkerchief'.DIM), we can reasonably assume that they have a shared representation 'small handkerchief', but if we use the word *maritino* ('hubby', lit. husband.DIM), it is much more difficult to establish a shared representation, if there is a shared representation at all, also because in this case the diminutive can have both a positive and ironic connotation. (Voghera 2023: 267)

The key notions that emerged from the outline presented in this section will be rearticulated in more detail in the following sections, along with the description and commentary of the data.

5. The corpus

The analysis is based on two different sets of data:

a. Data derived from a selection of the ethnotexts of ALS (*Atlante Linguistico della Sicilia*)[4], which are more or less extensive texts whose production has been stimulated by

---

[4] Most of the ethnotexts, collected for various research purposes, have been published in the approximatively 50 volumes of "Materiali e ricerche dell'ALS" ('ALS Materials and Research'), edited by the *Centro di studi filologici e linguistici siciliani* (CSFLS). Cf. Sottile (2019) for a comprehensive description of the ALS project. For this study, we selected the texts contained in ALS 1, 2, 32, 36, 37, which are among the quantitatively largest and qualitatively richest.

interviewers' questions (therefore it is semi-spontaneous) and have a largely monological and narrative character, as the speaker recounts or describes facts, experiences or traditions (Paternostro & Sottile 2010: 603).

b. Data derived from a questionnaire administered to 115 individuals (from different areas of Sicily, different ages and educational backgrounds). The questionnaire consists of 26 translation input questions (Italian to Sicilian, Sicilian to Italian), largely formulated on the basis of our native intuition. Input questions propose nouns, adjectives, adverbial formations, accompanied by some attenuative forms for translations from Italian into Sicilian (diminutive suffixes, periphrases, indefinite markers, prefixes etc.) and constructed with the suffix -*ḍḍu* in those from Sicilian into Italian[5].

The data obtained from translation questions, in which the input sentence to some extent determines the presence of a given form, are qualitatively different from those collected from ethnotexts. While the latter allow us to record the presence of a form in the informants' semi-spontaneous production, in translation questions we can observe the extent to which it is a conscious response. In this way, we hoped to activate different aspects of the speakers' competence and thus obtain more varied data[6].

For both datasets, we adopted a generic (and preliminary) *pan-Sicilian* approach, that is, we tried to draw a picture that, while representing different areas of Sicily, could provide

---

[5] However, we proposed especially nouns and this is reflected in the data obtained:

1. (Ital. > Sic.) 7 Ns (*discorsetto* '(serious) discourse, speech', *più o meno un'ora* 'about one hour', *freschetto* 'a little cool', *acqua marroncina* 'brownish water', *muratore di modeste capacità* 'mason of modest ability', *ramanzina* 'telling off', *fatterelli* 'facts of little importance, anecdotes'); (Sic. > Ital.) 7 Ns (*discurseḍḍu* '(serious) discourse, speech', *fattareḍḍu* 'fact of little importance, anecdotes', *mastriceḍḍu* 'mason of modest ability', *uriceḍḍa* 'little hour, about one hour', *acquiceḍḍa* 'sort of water', *cauliceḍḍi* 'type of wild edible vegetable').

2. (Ital. > Sic.) 5 ADJs (*un tantino difettoso* 'a little defective,' *semi-infossati* 'sunken', *un po' ammaccati* 'beated up', *tipo scimunito* 'kind of idiot', *leggermente ammaccato* 'slightly beated up'); (Sic. > Ital.) 2 ADJs (*scimuniteḍḍu* 'kind of idiot', *svinturateḍḍu* 'a little hapless').

3. (Ital. > Sic.) 3 ADVs (*tardi* 'late', *un po' tardi* 'a little late', *per circa un'ora* 'for about an hour'); (Sic. > Ital.) 2 ADVs (*pristuliḍḍu* 'a little early, quite early', *ppi faureḍḍu* '(little) please').

[6] An anonymous reviewer notes that translation fosters a strong impact in terms of calquing from Italian. This is undoubtedly plausible. However, it must be borne in mind that the interference between Sicilian and Italian goes far beyond the explicit translation tasks: it is a normal phenomenon for speakers who are bilingual, like our informants and like all speakers of Sicilian by now. In everyday linguistic practice, Italian and Sicilian alternate according to the context and, not infrequently, mix in the same discourse context. The choice to work simultaneously with (semi-)spontaneous data and with data that, on the other hand, reveal a metalinguistic awareness, such as translation tasks, reflects this pluri-linguistic situation and is inspired by a deep-rooted tradition of field research such as that of *Centro di studi filologici e linguistici siciliani* and the *Atlante Linguistico della Sicilia* (cf. D'Agostino and Ruffino 2005, 2013; Sottile 2019).

information on what is recurrent and systematic in the use of the suffix and allows us to leave aside for the moment (micro)diatopic issues[7].

Mainly because only a limited sample of the ALS corpus was used (especially in the case of ethnotexts), the description of the data carried out in the next sections will be exclusively qualitative[8].

6. Data analysis

The suffix forms mainly nouns (250 out of 262 suffixed words in *-ḍḍu*). Adjectives are much rarer and even fewer the other classes (i.e., markers of indefinite quantity, e.g., 'a little'). At least in part, this distribution might depend on the characteristics of the corpus itself, in particular the textual types represented especially in the ethnotexts, where suffixed nouns largely belong to specialised lexicons (sea, mining etc., according to the research conducted). However, one cannot ignore the existence of what Merlini Barbaresi (2004: 267) described as a "universal preference" in the selection of the basic category by the evaluative suffixes, with the noun in first place, followed by adjective, verb, adverb, indefinite, exclamation and finally numeral.

From a formal point of view, nouns are almost always derived from nominal bases (e.g., *sciccarieḍḍu*$_N$ 'little donkey' < *sceccu*$_N$ 'donkey'), with a few exceptions, namely, a few deverbal suffixed nouns, such as *cuçeḍḍa* 'steel needle for sewing creels' < *cusiri* 'to sew' and *tummar(i)eḍḍu* 'seagull/dolphin/tadpole' < *tummari* 'to fall head over heels, to dive', and a few deadjectival nouns, such as *russuliḍḍa*$_N$ < *russu*$_A$ 'red', *russeḍḍu*$_N$ < *russu*$_A$, *stuppateḍḍi*$_N$ < *stuppatu*$_A$ 'clogged', all designating specimens of marine fauna. However, these cases should be considered as lexicalised forms rather than productively constructed forms in speech (see Section 6.3). On the other hand, adjectives are all derived from adjectives, and adverbs from adverbs.

---

[7] This is particularly relevant as Sicilian lacks a standard variety. The data collected account for varieties from all the Sicilian provinces (Palermo, Catania, Agrigento, Caltanissetta, Enna, Messina, Ragusa, Siracusa, Trapani), but without any attempt to be systematic or balanced. The diatopic (micro)variation issue will deserve specific attention in the near future.

[8] In addition to the impossibility of quantitative generalisations, which only an analysis of the entire ALS database could legitimise, other considerations prompt a preference for the qualitative approach, that is, the textual specificities, which affect the types of data that emerge, and the lack to date of a micro-diatopic differentiation that could reveal differences in distribution of the suffix.

From a semantic point of view, corpus data can be described on the basis of Fig. 2. This is a way of schematically representing the semantics of -*ḍḍu*, which is inherently relational because it implies the existence of another reference entity that is unmarked and unmodified by the suffix. We describe this relational semantics in terms of prototypicality. The figure shows category X (e.g., 'donkey'), enclosed in the larger circle, and its members, enclosed in smaller inner circles. These members exhibit categorial properties to varying degrees, from the highest degree corresponding to the prototypical exemplar (X_1, the non-suffixed base, which is the centre of the category, e.g., *sceccu*$_N$ 'donkey') to the lower degrees represented by the more peripheral members (the possible meanings expressed by the derived form in -*ḍḍu*, e.g., 'little donkey', 'manger donkey' or even 'dunce' or 'jackass'). Within the larger circle, all members other than X_1 are equally referred to as X_2, as they coexist synchronically in the semantic periphery of X as non-prototypical members. We will refer to this semantic periphery in terms of "internal approximation". This means that the X_2 members only *approximate* to the centre of the category, i.e., they embody the categorial properties in an imprecise or attenuated form (which can in turn be instantiated as a diminished or non-prototypical or highly specific form, or a form marked pragmatically in an affective or ironic sense, etc.). Internal approximation (or attenuation) is, in line with Grandi's (2017) analysis, a specification of the higher-order semantic category "approximation"[9]. The adjective "internal" refers to the fact that the deviation from or non-coincidence with the categorial centre does not, however, exceed the boundaries of the category itself (thus a *sciccarieḍḍu*$_N$ is a non-central example of *sceccu*, but still a donkey; in other words, it is a hyponym for 'donkey').

Moving towards the left-hand side of Fig. 2, we find examples (X_3) that only *approximate to the category* as a whole, that is, they resemble the category but do not in fact belong to it or at least, their actual belonging to it is not clear-cut. We refer to this situation as "external approximation". This is another specification of the higher-order category approximation, where, unlike X_2, the relationship is between outside and inside the category. In other

---

[9] The term approximation suggests a 'movement towards', but it is worth emphasising that the use of the term here is neutral in this sense, as indicated by the bidirectionality of the arrows in Fig. 2. Here centripetal orientation reflects the notion of approaching predefined values by members that do not fully meet those values; in contrast, centrifugal orientation expresses *dis*-proximation, that is, deviation from default values (see also Cappelle, Daugs & Hartmann 2023).

words, the presence of the suffix has an impact on the categorial membership: it does not produce hyponyms of X, but indicates exemplars that resemble X but are distinct from it ('similar to X', 'resembling X', 'sort of X'). This can occur at different levels: X-*ḍḍu* shows such highly specific characteristics as ends up being intersubjectively recognised as autonomous from X (e.g., *figureḍḍa* 'holy card' < *figura* 'image' discussed in Section 1; see also the already mentioned case of *fazzolettino* 'kleenex' < *fazzoletto* 'handkerchief' reported by Voghera 2023, where the attenuated form is somehow conventionalised as a new meaning, while easily still traceable to the meaning of the base). A true case of semantic autonomy from the base, such that X-*ḍḍu* is only related to X by a similarity relationship, is for example *lattucheḍḍa* 'seaweed', which is not a hyponym of (i.e., a 'type of') *lattuca* 'lettuce', although described in relation to this. However, even in this case, membership is largely dependent on the perspective of the speaker (S, in Fig. 2), which may be highly subjective (and possibly induced by specific discursive coordinates, as in the case of the attenuated forms X_2) or shared, as in the case of X_3, within a cultural-linguistic community (i.e., *intersubjective*, in the terms used by Voghera 2023).

Outside the area of subjectivity, on both the left and the right side, we find examples (X_n) for which the sense of (internal or external) approximation has been lost. Here the suffix can be considered lexicalised and the semantic relationship between the X-*ḍḍu* words and the bases is opaque. For instance, *vaccareḍḍa* '(edible) snail' is based on *vacca* 'cow', but it does not refer to a non-prototypical example of a cow nor an entity resembling a cow; it is something that at some point, for reasons no longer clear, was described in terms of a cow and thus crystallised in the lexicon (cf. Lanaia 2013).[10]

Following the schema in Fig. 2, from the inside to the outside, the different semantic outputs of suffixation with -*ḍḍu*/*a* will be described in Sections 6.1 to 6.3.

---

[10] Lexicalised forms of this type (not limited to the present case) may or may not share one or more characteristic traits with the base (e.g., in the case of *vaccareḍḍa*/*vacca* we could perhaps think of horns), but in other cases they are united not by semantic but textual factors, e.g., by childish nursery rhymes (Lanaia 2013: 86). However, these are relations that can only be reconstructed.
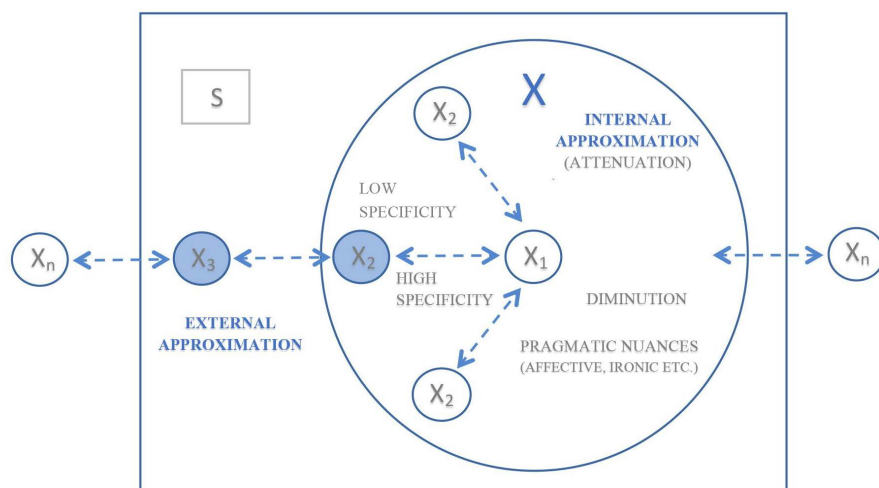
Fig. 2: Degrees of approximation expressed by *-ḍḍu/ḍḍa*

6.1 Internal approximation or attenuation (X$_2$)

In a substantial group of words in the ethnotexts, *-ḍḍu* has attenuative value (X$_2$). This value can be specified as diminution. In this case, the derivative designates a specimen that is smaller (or younger) than that denoted by the base, as *faragghiuneḍḍi* (< *faragghiuni* 'sea stack') in (3) and *picciutteḍḍu* (< *picciottu* 'boy') in (4):

(3)   *All'ùttimu poi cci su ḍḍi quaṭṭru scogghi ca di ccà mancu si vèdonu. Si chiàmanu* **faragghiuneḍḍi** [ALS 37, 113]

'At the end there are those four rocks that you can't even see from here. They are called **little sea stacks**'

(4)   *Io era* **picciutteḍḍu** *ma a mmìa m'aṛṛistaru tutto imppresso chiḍḍu chi façìanu iḍḍi* [ALS 32, 121]

'I was a **little boy** but the things they did all stuck in my mind'

In many cases, attenuation is associated with a set of pragmatic values pertaining to affective attitude, such as endearment, euphemism, irony, politeness. This means that the suffix expressing such meanings has scope on the entire speech act and work as *pragmatic index* that modulates the tone of the utterance (Brucale & Mocciaro 2020: 65). Cases in point are the nouns *puninteḍḍu* 'light westerly breeze' (< *punenti* 'westerly breeze') and *vizzicieḍḍu* 'little vice' (< *vizziu* 'vice'), in (5) and (6):

(5)   *u **punineḍḍu** u ccumpagnava comu fidili cumpagnu di sotta e di svintura* [ALS 37, 490]

'The **light westerly breeze** accompanied him as a faithful companion of fate and misfortune'

(6)   *A Ccisarò s'ammàzzunu a urpi e ss'a màncinu, cci l'annu ssu **vizzicieḍḍu*** [ALS 32, 211]

'In Cesarò they kill foxes and eat them: they have this **little vice**'

In these cases, the suffix operates at the level of the utterance as a mitigating device (see Caffi 2007). In particular, in (5), the attenuated derivative functions as "an intimacy accelerator" and empathetically brings the listener closer to the content of the utterance (Brown & Levinson 1987: 103). On the other hand, in (6) the suffix is used to mitigate (possibly ironically) an overly sharp categorical attribution. Indeed, 'vice' inherently conveys a negative meaning and would be placed in a taboo context (i.e., fox-hunting). By using the attenuated form, the speaker makes her assessment of the event explicit by describing it as a marginal or minor case of vice.

Within the same domain of attenuation, we can place the few adjectival and adverbial formations in *-ḍḍu* in the corpus. The former are gradable adjectives or departicipial formations expressing attenuation of the quantity or the positive or negative quality expressed by the base (cf. Emmi 2011: 237, based on Merlini Barbaresi 2004), e.g., *lungareḍḍa* 'a bit long' < *longa* 'long', *grossuliḍḍi* 'a bit thick'< *grossi* 'thick', *ammaccateḍḍu* 'slightly dented' < *ammaccatu* 'dented'. Also in this case, attenuation may result in a pragmatic modulation of the general tone of the utterance, as in (7):

(7)   *Giovanni è **scimuniteddu** ma è un travagghiaturi* [from the questionnaires]

'Giovanni is **a bit dumb,** but he is a hard worker'

As in the case in (6), the speaker is making an assessment, more specifically giving a negative evaluation of a person. By choosing the attenuated form, the speaker weakens the predictable unpleasant effects of the speech act and, thus, shows empathy because s/he is able to calculate the implications of what is being said (Caffi 2007: 82–83).

Adverbs play similar functions. They are all derived from generic quantifiers, e.g., *(t)anticchieḍḍa* 'a little' < *(t)anticchia* 'a little' and *cchjessanieḍḍu* 'a little more' < *cchiossai* 'more'. These forms attenuate the meaning of the (mainly adjectival) base (which may itself be suffixed with *-ḍḍu*), possibly adding a pragmatic specification. This is the case in (8),

where the suffix is used with both the adverb and the adjective to attenuate the possible impoliteness of the utterance:

(8)   *Avi l'ùocchi* **anticchìeḍḍa scavatieḍḍi** *ma è bìeḍḍa u stissu* [from the questionnaires]
      'She's a bit sunken-eyed but she is still beautiful'

If in (6)–(8) the negative sense is contained in the lexical meaning of the bases ('vice', 'dumb', 'sunken') and is attenuated by the suffix. Other times it is instead the suffix itself that suggests a negative interpretation of the situation described. This can be observed in (9), where the suffix is used to attenuate not the tone of the utterance but the categorial properties of the base, so that a *mastriceḍḍu* is a mason with poor skills, a mediocre mason (i.e., not what a prototypical mason should be):[11]

(9)   *Pino è* **mastriceḍḍu** *ma a picca a picca...* [from the questionnaires]
      'Pino is a **mediocre mason**, but little by little... (he is improving)'

Pragmatic attenuation can also be observed in what can be called 'just' or 'merely' meaning[12]. This is shown in (10), where the suffix does not attenuate or alter the categorial properties of the base, since *uovicieḍḍu* (< (*u*)*ovu* 'egg') is not a little or non-prototypical egg, but simply an egg. What the suffix attenuates here is the importance of the act recounted by the speaker (who barely ate an egg). The speaker here intends to make the hearer grasp the general meaning of the utterance and *uovicieḍḍu* is to be intended as a pure example of frugal meal in general[13]:

---

[11] Elsewhere in the corpus, the assessment is more overtly negative by virtue of the presence of other contextual elements that reinforce this interpretation, as in the case of the adjective *scarsu* 'poor, mediocre' in *Pinu jè un mastriciḍḍu scarsu ma sta addivintannu cchiù migliu* 'Pino is a poor mason, but he is improving.' Here the use of the attenuated derivative is actually more similar to that in (6) and (7), where it is employed to make the assessment less sharp and more acceptable and empathetic. Context always plays a crucial role in the interpretation of these nuances.

[12] Jurafsky (1996: 556–557) explains this type of meaning as a declination on the pragmatic level of what he calls 'lambda-abstraction', i.e., the specific mechanism of attenuated sense construction by virtue of which second order predicates develop, marking approximation. Among the operators that induce lambda-abstraction, Jurafsky lists expressions such as loosely speaking, only, just, merely etc., which contain a second speech act, i.e., a metalinguistic comment that serves to modulate the illocutionary force of the utterance and to limit its scope.

[13] Of particular note is the co-occurrence with the element *panuzzu* 'bread', itself characterised by the attenuative suffix *-uzzu*, and with a quantifier of a mass noun, *um-muccuni* 'a morsel', both emphasising quantitative scarcity.

(10)   *Siempri ḍḍà simu: panuzzu sciuttu, n **uovicieḍḍu**, chiḍḍu era u manciari. L'uvi-cieḍḍu e um-muccuni i vinu e ffinisci ḍḍuocu* [ALS 36, 202]

'Always the same: dry bread, **barely an egg**, that was the meal. Barely an egg and a sip of wine and that's it'

Sometimes, rather than producing pragmatic mitigation effects as observed, with various nuances, in (5) to (10), the suffix may serve to underspecify the meaning of the base, expressing a value approximating (by default or excess) the prototypical value of the category. This is the case of *uriceḍḍa* (also *urateḍḍa*) < *ura* 'hour', in (11):

(11)   *Vignu nni tia 'nt'a n'**uriceḍḍa*** [from the questionnaires]
'I'll come to you in an **hour or so**'

Here *uriceḍḍa* (like the already mentioned *chiliceḍḍu*, see Section 1) does not indicate 'a little hour' or 'less than an hour', but rather 'approximately an hour (maybe a little less, maybe a little more)'. In other words, it is an inaccurate, therefore not prototypical specimen, of 'hour'.

Quite the opposite, attenuation may end up developing meanings (intersubjectively) perceived as denoting specific exemplars of the category. This can be seen by comparing the use of *mastriceḍḍu* in (12) with that already observed in (9):

(12)   *Pinu è un mastriciḍḍu ma sa fida* [from the questionnaires]
'Pino is a basic/non specialised mason, but is capable'

Unlike (9), here *mastriceḍḍu* simply describes the competence of the referent, the scope of his specialisation, without conveying a negative sense (which would be contradictory: [?]'Pino is scarce but skilful'); in other words, *mastriceḍḍu* denotes a specific type of *mastru* 'mason'. A similar case is *discurseḍḍu* (< *discursu* 'discourse, argument, speech'), which is not 'a little discourse' but always 'a scolding (or at least a speech with serious content)', without any reference to its duration, as exemplified in (13):

(13)   *Ci fici un discurseḍḍu e mi dissi ca unnu fa cchiù* [from the questionnaires]
'I gave him a telling off and he said he won't do it again'

While in the case of *mastriceḍḍu* the specific nuance remains context-dependent, as the cases in (9) and (12) (and in fn. 11) show, in (13) the value of 'high specificity' conveyed by

the suffix is more firmly associated with the derived form *discurseḍḍu*. We will come back to this example in 6.2.[14]

6.2 From internal to external approximation (X₃)

The case of *discurseḍḍu* in (13) is situated in an area of transition to another semantic sub-domain, in which the suffix does not attenuate the categorial properties of the entity denoted by the base, as in the cases in 6.1, but challenges them. In other words, the exemplar encoded by the suffixed word is perceived as so specific or non-prototypical that it may end up being considered something else, outside the category. These are the cases represented as X₃ on the left-hand side of Fig. 2: such specimens are external to the category X and, in fact, do not approximate the centre of X, but the category as a whole. A clear example of X₃ is *figureḍḍi* 'holy cards' in (14):

(14)   *Nn'avìvavu santi a cui tinìvavu particolarmenti? Santi protettori, **figureḍḍi**...* [ALS 32, 91]
         'But didn't they have saints they particularly cared about? Patron saints, **holy cards**...'

Similar to the case of *fazzolettino* reported by Voghera (2023) (see Section 4), we can assume an "intersubjectively shared representation" for *figureḍḍi* in (14), whose membership to the category 'figure' cannot be said to be clear-cut. In fact, *figureḍḍi* is still a type of figure, but so specific that it tends to embody an autonomous entity (with recurring properties that make it partly different from other figures, e.g., small size, the type of subjects represented, the context of use etc.)[15].

An entity independent of the one denoted by the base is certainly that expressed by *lattucheḍḍa* with the meaning 'seaweed' (e.g., ALS 37: 462), which originates from a (metaphorical) extension of the base *lattuca* 'lettuce': a *lattucheḍḍa* is 'a sort of lettuce', 'an entity resembling a lettuce', but it is not a lettuce. The speakers are aware of the similarity relation

---

[14] An anonymous reviewer observed that an implication of smallness would be maintained here. However, it is actually the attenuative function that remains in the background, which is used to downplay, for ironic or sarcastic purposes, the strength of the utterance. In any case, *discurseḍḍu* is to be considered broadly (intersubjectively) conventionalised in its sense of 'serious discourse'.

[15] The corresponding Italian words *immaginetta* (< *immagine* 'image' + evaluative feminine suffix -*etta*) and *santino* (< *santo* 'saint' + evaluative suffix -*ino*) function in the same way. Another relevant example, which may help interpret X₃ cases, is the Italian *figurina* in its meaning of 'sticker', that is, a small colour image, typically sold in sachets, of which one makes a collection'. Another transparent case is Sic. *bambineḍḍu* 'child' + evaluative, hence par excellence 'Christ child' (cf. Italian *bambinello* 'child' + evaluative suffix, with the same meaning). Cf. Jurafsky (1996: 552), who described similar lexical drifts for various languages, including non-Indo-European ones.

(the referent is also known as *lattuga marina* 'sea lettuce'; ALS 37: 463) and use this similarity to name a referent as 'sort of X' or 'X-like', which is a full example of "external approximation"[16].

The slightly different readings of the types *figureḍḍa* and *lattucheḍḍa* compared to the cases discussed in 6.1. lies in the shift from a purely subjective dimension of evaluation (whereby the individual speaker decides on a case-by-case basis whether to use a suffix to express specific nuances of meaning) to an intersubjective one (on some level already present in *discurseḍḍu*, but here unequivocal), which is anchored to representations that are likely to be shared by a community of speakers. Although qualitatively different, the examples X_2 in 6.1 and the two types X_3 are linked by a relation of continuity, as they express an increasing distancing (indicated by the dashed arrows in Fig. 2) from the central properties of the category X and then from the category itself (with several uncertain cases at the boundaries, as is normal in the prototypical approach), which in any case remains a transparent point of reference for the interpretation of the suffixed word. In other words, the suffixed words of the types X_3 remain fully compositionally transparent. Precisely because of this transparency, we include in this area the numerous metaphorisations attested in the corpus, e.g., words denoting different specimens of marine fauna by referring to bases with the trait [-animate], such as *ancileḍḍru* (also *anciuliḍḍru*) 'owl fish' (ALS 37: 369) < *anci(u)lu* 'angel' (the fish has large pectoral fins resembling wings).

However, since examples X_3 tend to undergo lexical drift, they may overstep the boundaries of intersubjectivity and end up conveying meanings no longer referable to the base at the synchronic level. These cases, which are placed outside the area in which S has its scope, are therefore classified in the group of lexicalisations that will be discussed in 6.3.

---

[16] Of course, elsewhere *lattucheḍḍa* (as well as the other derivatives that have conventionalised specific meanings) may retain an effectively attenuative value of 'lettuce', 'lettuce sprout' etc., alongside other specific meanings such as 'wild lettuce' or even entirely autonomous ones such as 'specific cut of beef', recorded in Piccitto & Tropea (1985: II, 458). This depends very much on the variety of Sicilian observed and, in any case, the co-presence of several meanings cannot be surprising in a prototypical view of semantic categories.

6.3 From EA to the loss of transparency (Xₙ)

Many nouns in -ḍḍu found in the ethnotexts are lexicalised formations at different levels of opacity, i.e., the relationship between the base and the suffix is no longer transparent from a synchronic point of view. We can identify different cases, among which:

1. Suffixed nouns with a different gender than the base (masculine to feminine or feminine to masculine). The change of gender (a fact already observed for Sicilian, cf. Emmi 2011: 234–235) is often accompanied by a change of referent; for instance, if *purteḍḍ(r)u* (m) 'hatchway' (ALS 37: 191) could still be analysed as a highly specific type of *porta* (f) 'door', *calamareḍḍa* (f) 'squid fishing bait' (ALS 37: 286) exhibits only a generic relationship with the base *calamaru* (m) 'squid' (something related to squids).

2. Formations that change the lexical category of the base, such as *cuçeḍḍa*ₙ 'steel needle for sewing creels' < *cusiri*ᵥ 'to sew', *tummar(i)eḍḍu*ₙ 'gull, dolphin, tadpole' < *tummari*ᵥ 'going underwater while already at sea'; *ṛṛusseḍḍu*ₙ 'red mullet or strawfish whitebait' (ALS 37: 326) < *russu*ₐ𝒹𝒿 'red'.

3. Non-transparent formations. At least in principle, the suffix -ḍḍu could be formally segmented, but its semantic contribution is not clear, as in *vardeḍḍa* 'packlesaddle or saddlebag' (< *varda* with the same meaning) or went lost. In many cases, in fact, the process of word formation can only be reconstructed by going back to different phases of the language (cf. *picciriḍḍu* or *picciliḍḍu* 'small, child', based on the root **pikk* 'tip', hence 'small', probably through a now obsolete *picciulu* 'small'[17]).

In all these cases, we are outside the domain of subjectivity, hence of the speaker's evaluation (S). The relationship between the suffixed forms and the bases from which they derive is opaque, or it is not possible to recognise the referent of the base; therefore, the meaning

---

[17] In the early Sicilian texts collected in ARTESIA, we already find *picculillu*ₐ𝒹𝒿 'small, irrelevant' (*MuntiXVC*, 8.32.1), *picchulilli*ₙ 'children' (*QuaedProfXIVC*, 21.24.5), *pichirillu*ₐ𝒹𝒿 'small' (*ThesaurusXVR*, 155.5, 82.16; *MeditacioniXVGQ,* 9.41.4) and *pichirillu*ₙ (*MeditacioniXVGQ,* 9.40.25). According to TLIO (*Tesoro della lingua italiana delle origini*, http://tlio.ovi.cnr.it/TLIO/; the etymology is based on DEI), *pichirillu* is to be traced back to *picciolo* 'small'. However, Sic. *picciulu* 'small' is rare and no longer transparent for most varieties, where the adjective *nicu* 'small' prevails. *Picciulu* (as well as *picciottu* 'boy', see Varvaro 2014: 759) < **pikk* 'small' (Meyer-Lübke 1911: 483) or 'tip' hence 'small' (cf. for instance Battaglia 1986: 352, 355, 364). According to Di Gregorio (1936: 253), *picciriḍḍu* is to be traced back directly to this root. In both the reconstructions, the base is opaque.

of the -*ḍḍu* form is totally memorised (significantly, these words are usually lemmatised in lexicons or vocabularies, as opposed to those productively derived; cf. also Emmi 2011: 238). Even when it is possible to identify a semantic relationship between base and suffixed word, it is still a generic relationship quite different from approximation (an entity somehow connected to the meaning of the base, e.g., having the colour indicated by the base or doing what the base indicates etc.).

7. (Provisional) conclusions

The analysis conducted in Sections 6.1 to 6.3 was based on the representation of approximation proposed in Fig. 2 in terms of prototypicality and allowed us to isolate three groups of derivatives in which the function of the suffix -*ḍḍu*/*ḍḍa* can be observed. The first group (X_2), which is also the most numerous, contains productively derived words in which the suffix plays an attenuative function, that is, it indicates that the referent of the derived word exhibits the categorial properties of the base to a reduced degree. Attenuation is instantiated variously, depending on the context, as diminution (*sacchitieḍḍu* 'small bag' < *saccu* 'bag', *fimminieḍḍa* 'small exemplar of woman, young woman' < *fimmina* 'female, woman'), mitigation for the purposes of affection or empathy (*mischineḍḍu* < *mischinu* 'unfortunate'), courtesy (e.g., *scimuniteḍḍu* 'a little dumb'), etc. Sometimes the attenuation of categorial properties may result in under-specification (e.g., *uriceḍḍa* 'about one hour' < *ura* 'hour') or in a reduction in the extension of the meaning of the base X, such that the suffixed word expresses a specific exemplar of X, the semantic representation of which is likely to be intersubjectively shared (e.g., *discurseḍḍu* 'telling-off, harsh argument' < *discursu* 'speech, discourse'). This latter case marks the transition to external approximation (X_3), i.e., an area that contains exemplars that are no longer hyponyms of the base, because their specificity is, in intersubjective evaluation, so pronounced as to denote autonomous entities distinct from X rather than non-prototypical examples of X (*figureḍḍa*) or entities which only resemble X (*lattucheḍḍa*). Beyond this area, on the left and right peripheries of the schema, we find the (rather numerous) examples in which the suffix has lost any form of productivity and transparency and is, instead, fully lexicalised. Even when the suffix is formally identifiable, the relationship with the base is only generic (e.g., *calamareḍḍa*

'squid fishing bait', *ṛṛuseḍḍu* 'red mullet') or the base is not recognisable (e.g., *picciriḍḍu* 'child'). Examples X_n, which presumably survive from other phases of the language, are therefore beyond the (evaluative, hence subjective) domain in which approximation operates and, in fact, are placed in the scheme in Fig. 2 outside the box marked S(ubject).

Based on this classification, we can extract a semantic-functional map for the suffix *-ḍḍu/-a*, as represented in Fig. 3. In the map, the various meanings attested in the corpus (in grey) can be arranged around two semantic cores (in blue), that is, 'X_2, attenuation' (internal approximation) and 'X_3, similar to' (external approximation). The map does not include the examples X_n, which are not derived by means of a productive use of the suffix, that is, the role of the suffix is not transparent in these cases.
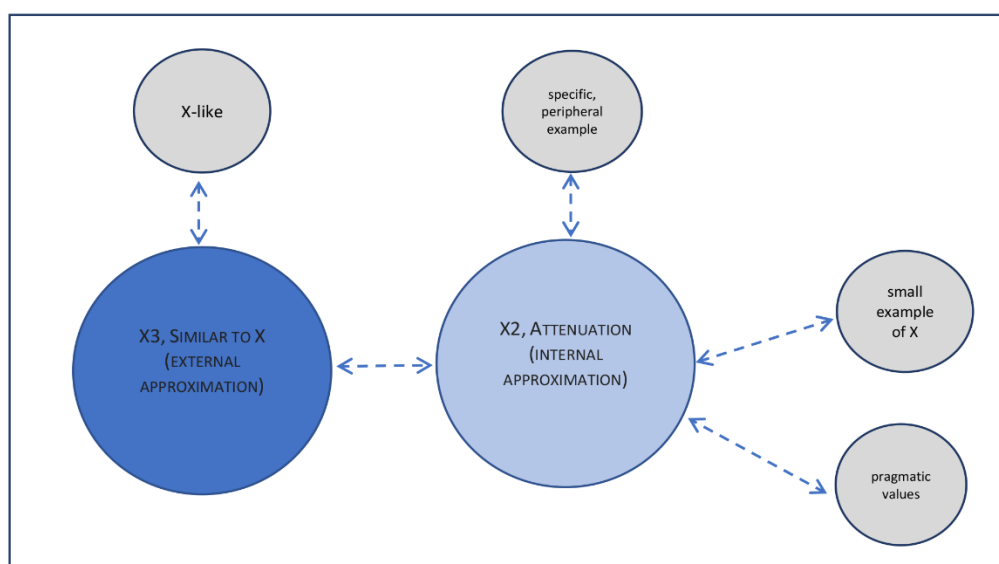
Fig. 3: Towards a semantic map for *-ḍḍu/-a*

Several aspects of the analysis deserve further investigation in the near future, which will allow a more comprehensive description of the range of meanings of the Sicilian suffix *-ḍḍu/-a*. Among these:

1. A systematic diachronic analysis of individual lexicalised words in order to assess whether other functions of the suffix can be traced as productive mechanisms in different phases of the language (e.g., the generic relation function observed for Latin and only residual and not very transparent in contemporary Sicilian).

2. The evaluation of possible local differences, to be conducted by means of new questionnaires targeted at different points in Sicily.

3. Competition with other evaluative (diminutive) suffixes in the expression of approximation (e.g., *-ignu* '-ish', e.g., *umitignu* 'dampish' < *umitu* 'damp') or with different means (e.g., generic markers of quantity, among which indefinite adverbial items *tanticchia*, *nna picca* 'a bit', 'half items', e.g., *mezzu* 'half', etc., or other means, e.g., *tipu* 'type', as emerges from the responses to the questionnaire).

This is food for further research.

References

ALS = *Atlante Linguistico della Sicilia* [Linguistic Atlas of Sicily]. http://atlantelinguisticosicilia.it (accessed 13 June 2022).

ALS 2 = Ruffino, Giovanni. 1995. *I pani di Pasqua in Sicilia. Un saggio di geografia linguistica e etnografica* [Easter Breads in Sicily. An Essay on Linguistic and Ethnographic Geography]. Palermo: CSFLS (Materiali e ricerche dell'ALS 2).

ALS 32 = Castiglione, Marina. 2012. *Parole e strumenti dei gessai in Sicilia. Lessico di un mestiere scomparso* [Words and Tools of the Chalk Workers in Sicily. Lexicon of a Vanished Trade]. Palermo: CSFLS (Materiali e ricerche dell'ALS 32).

ALS 36 = Castiglione, Angela. 2016. *Le parole del cibo. Lingua e cultura dell'alimentazione a Troina* [The Food Words. Language and Culture of Food in Troina]. Palermo: CSFLS (Materiali e ricerche dell'ALS 36).

ALS 37 = D'Avenia, Elena. 2018. *Atlante linguistico della Sicilia. Il lessico del mare* [Linguistic Atlas of Sicily. The Sea Lexicon]. Palermo: CSFLS (Materiali e ricerche dell'ALS 37).

ARTESIA = *Corpus Artesia. Archivio testuale del siciliano antico* [Artesia Corpus. Textual Archive of Old Sicilian]. http://artesia.ovi.cnr.it/ (accessed 13 June 2022).

Battaglia, Salvatore. 1986. Grande dizionario della lingua italiana [Large Dictionary of Italian Language]. Vol. XIII. Torino: Utet.

Brown, Penelope & Stephen C. Levinson. 1987. *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press.

Brucale, Luisa & Egle Mocciaro. 2020. Suffissi valutativi in Plauto: una proposta cognitivista e morfopragmatica [Evaluative Suffixes in Plautus: A Cognitivist and Morpho-Pragmatic Account]. *Lingue antiche e moderne* 9. 57–86. DOI: 10.4424/lam92020-3.

Brugmann, Karl. 1906. *Grundriß der vergleichenden Grammatik der indogermanischen Sprachen, vol. II/1*, II edition. Straßburg: Trübner.

Caffi, Claudia. 2007. *Mitigation*. Amsterdam & Tokio: Elsevier.

Cappelle Bert, Robert Daugs & Stefan Hartmann. 2023. The English privative prefixes *near-*, *pseudo-*, *quasi-* and *sub-*: Approximation and 'disproximation'. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1). 52–79.

D'Agostino, Mari & Giovanni Ruffino. 2013. L'esperienza dell'Atlante Linguistico della Sicilia [The Experience of the Linguistic Atlas of Sicily]. In Giovanni Ruffino (ed.), *Lingue e culture in Sicilia*

[Languages and Cultures in Sicily], 1434–1454. Palermo: Centro di studi filologici e linguistici siciliani.

D'Agostino, Mari & Giovanni Ruffino. 2005. *I rilevamenti sociovariazionali. Le linee progettuali* [The Socio-Variational Surveys. Project Lines]. Palermo: Centro di studi filologici e linguistici siciliani.

De Gregorio, Giacomo. 1907. Suffissi di significato diminutivo in siciliano [Suffixes with Diminutive Meaning in Sicilian]. *Studi glottologici italiani* 4. 211–238.

DEI = Carlo Battisti & Giovanni Alessio, 1950–57. *Dizionario Etimologico Italiano* [Italian Etymological Dictionary]. Firenze: Barbèra.

Dressler, Wolfgang U. & Lavinia Merlini Barbaresi. 1994. *Morphopragmatics: Diminutives and Intensifiers in Italian, German, and Other Languages*. Berlin: De Gruyter.

Emmi, Tiziana. 2011. La formazione delle parole in siciliano [Word Formation in Sicilian]. Palermo: Centro di studi filologici e linguistici siciliani.

Grandi, Nicola. 2017. I diminutivi come marche di attenuazione e indeterminatezza [Diminutives as Markers of Attenuation and Indeterminacy]. In Oana-Dana Balaş, Adriana Ciama, Mihai Enăchescu, Anamaria Gebăilă & Roxana Voicu (eds.), *L'expression de l'imprecision dans les langues romanes*, 162–175. Bucharest: Ars docendi – Universitatea din Bucureşti.

Grandi, Nicola. 2015. Intensification processes in Italian. A survey. In Maria Napoli & Miriam Ravetto (eds.), *Exploring Intensification. Synchronic, Diachronic, and Cross-linguistic Perspectives*, 55–77. Amsterdam: Benjamins.

Grandi, Nicola. 2011. Renewal and innovation in the emergence of Indo-European evaluative morphology. *Lexis 6: Diminutives and Augmentatives in the Languages of the World.* 5–25. https://journals.openedition.org/lexis/403 (accessed 26 June 2022).

Grandi, Nicola & Lívia Körtvélyessy (eds). 2015. *Edinburgh Handbook of Evaluative Morphology*. Edinburgh: Edinburgh University Press.

Hakamies, Reino. 1951. *Étude sur l'origine et l'évolution du diminutive latin et sa survie dans les langues romanes*, Helsinki: Suomalaisen Kirjallisuuden Seuran Kirjapainon Oy.

Hamawand, Zeki. 2017. The notion of approximation in language: A construal-based analysis. *Cognitive Semantics* 3(1). 95–120.

Haverling, Gerd. 2011. La suffixation 'diminutive' du latin préclassique et classique au latin tardif. In Michèle Fruyt & Olga Spevak (eds.), *La quantification en latin*, 225–257. Paris: L'Harmattan.

Jurafsky, Daniel. 1996. Universal tendencies in the semantics of the diminutive. *Language* 72(3). 533–578.

Lakoff, George. 1987. *Women, Fire, and Dangerous Things*. Chicago: University of Chicago Press.

Lanaia, Alfio. 2013. *Nomi siciliani di invertebrati e piccoli animali. Studio etimologico e iconimico* [Sicilian Nouns of Invertebrates and Small Animals. An Etymological and Iconymic Study]. Unpublished Doctoral thesis, University of Catania.

Magni, Elisabetta. 1999. Sémantique du diminutif et relations d'appartenance. *Silexicales* 2. 139–148.

Merlini Barbaresi, Lavinia. 2004. Alterazione [Alteration]. In Maria Grossmann & Franz Rainer (eds.), *La formazione delle parole in italiano* [Word Formation in Italian], 264–292. Tübingen: Niemeyer.

Meyer-Lübke, Wilhelm. 1911. *Romanisches Etymologisches Wörterbuch*. Heidelberg: Winter.

Mutz, Katrin. 2015. Evaluative morphology in a diachronic perspective. In Nicola Grandi & Lívia Körtvélyessy (eds.), *The Edinburgh Handbook of Evaluative Morphology*, 142–154. Edinburgh: Edinburgh University Press.

Paternostro, Giuseppe & Roberto Sottile. 2010. Parlante, identità e (etno)testo fra storia della lingua e dialettologia. L'esperienza dell'Atlante Linguistico della Sicilia [Speaker, Identity and (Ethno)text between Language History and Dialectology. The Experience of the Linguistic Atlas of Sicily]. In Giovanni Ruffino & Mari D'Agostino (eds.), *Storia della lingua e dialettologia* [Language History and Dialectology], 597–614. Palermo: Centro di studi filologici e linguistici siciliani.

Piccitto, Giorgio & Giovanni Tropea. 1985. *Vocabolario Siciliano* [Sicilian Dictionary]. Palermo: CSFLS.

Prieto, Victor M. 2015. The semantics of evaluative morphology. In Nicola Grandi & Lívia Körtvélyessy (eds.), *The Edinburgh Handbook of Evaluative Morphology*, 21–31. Edinburgh: Edinburgh University Press.

Rohlfs, Gerard. 1969. *Grammatica storica della lingua italiana e dei suoi dialetti. Sintassi e formazione delle parole* [Historical Grammar of the Italian Language and its Dialects. Syntax and Word Formation]. Torino: Einaudi.

Sottile, Roberto. 2019. L'Atlante Linguistico della Sicilia (ALS) [The Linguistic Atlas of Sicily (ALS)]. In Thomas Krefeld & Roland Bauer (eds.), *Korpus im Text* 7. http://www.kit.gwi.uni-muenchen.de/?p=34845&v=1 (accessed 26 June 2022).

Varvaro, Alberto. 1988. Italienisch: Areallinguistik XII. Sizilien. Aree linguistiche XII. Sicilia. In Günter Holtus, Michael Metzeltin & Christian Schmitt (eds.), *Lexicon der Romanistischen Linguistik. Bd. 4. Italienisch, Korsisch, Sardisch*, 716–731. Tübingen: Niemeyer.

Varvaro, Alberto. 2014. *Vocabolario storico-etimologico del siciliano* [Historical-Etymological Dictionary of Sicilian]. Palermo: CSFLS.

Voghera, Miriam. 2023. The role of diminutive suffixes in the Italian Time Nouns constructions: From approximation to focus? *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1). 263–286.

Luisa Brucale
Università di Palermo
Dipartimento Culture e Società
Viale delle Scienze, ed. 15, 90128 Palermo, Italy
luisa.brucale@unipa.it

Egle Mocciaro
Masaryk University
Faculty of Arts, Department of Romance Languages and Literatures
Arna Nováka 1/1, 602 00 Brno, Czech Republic
egle.mocciaro@mail.muni.cz

Bert Cappelle, Robert Daugs & Stefan Hartmann

The English privative prefixes *near-*, *pseudo-* and *quasi-*: Approximation and 'disproximation'

Abstract: The English prefixes *near-, pseudo-* and *quasi-* are privative, in that whatever essential property their morphological base expresses is not strictly possessed by an entity characterized as *near-/pseudo-/quasi-*X. However, we claim this meaning is not precise enough and hypothesize that *near-* and *quasi-* are approximative in meaning, whereas *pseudo-* is 'disproximative', expressing the idea of 'falling short' of a standard. Distributional-semantic findings partially support this, as *near-* shares more bases with *quasi-* than it does with *pseudo-*. *Near-* is most productive, presenting a default choice, while *pseudo-* is least productive. We also observe a specific tendency of *near-* to select bases with negative semantic prosody (*near-deadly, near-fatal*), of *quasi-* to combine, without any evaluative meaning, with legal-administrative bases (*quasi-diplomatic, quasi-governmental*), and of *pseudo-* with terms from the scientific domain. Further qualitative observations about these prefixes are made.

Keywords: approximation, disproximation, distributional semantics, prefix, privativity, productivity, semantic prosody

1. Introduction[*]

1.1 Privativity and approximation

This paper is concerned with three prefixes with a somewhat similar meaning: *near-*, *pseudo-* and *quasi-*, as in *near-perfect*, *pseudo-scientific* and *quasi-religious*. From a formal-semantic point of view, all three count as 'privative' prefixes. A privative morpheme, whether bound or free, has the effect of removing an essential property from an entity. Thus, *un-* in *unhappy* is privative in that it indicates the absence of happiness in whoever is described as *unhappy*. Likewise, the adjective *fake*, applied to for instance *blood*, is said

---

to be privative because the stuff that we call *fake blood* misses some essential properties of real blood, such as the ability to transport oxygen through the body.

Privativity, however, is too crude a concept to allow us to grasp the full meaning of operators lumped under this semantic heading, or even to characterize those operators consistently (Kamp 1975: 125; Boleda et al. 2012: 1228; Nayak et al. 2014: 5; Cappelle, Denis & Keller 2018). For example, both *non-* and *near-significant* can be applied to findings that are not significant, and both prefixes are therefore privative. Yet, privativity by itself does not help us here to capture the observation that a science project's *non*-significant results are, on the whole, further removed from any significant results than its *near*-significant ones are. As for non-consistency, consider again the adjective *fake*. While *fake blood* does not at all qualify as blood, a *fake handbag* still refers to a handbag, with full functionality – that it is not actually manufactured by a factory of the brand it carries does not remove any essential 'handbagginess' from the entity referred to as being fake. *Fake*, therefore, is not consistently privative across all its uses. As for *near-*, *pseudo-* and *quasi-*, these prefixes convey privativity consistently: for all lexical items that can fill the position of X, *near-/pseudo-/quasi-*X is never, in fact, truly or fully X.

While in our example of *non-* versus *near-* above, only *near-* could be said to be 'approximative' (in addition to being privative), all three of the prefixes focused on in this paper can be thus qualified, in a broad sense at least (see the introduction of this special issue for some background on 'approximation'). For instance, a *near-human*, *a pseudo-human* and a *quasi-human* refer to some entity that is not a human but that, in terms of looks or behaviour, approximates one; this can also be said of a *near-/pseudo-/quasi-human creature*, where the prefix is added to an adjectival base. So, technically, the referent of a noun with any of these prefixes, or of a noun combined with an adjective prefixed by any of them, can be interpreted as having properties in common with referents of nouns without any such prefix, or appearing with the corresponding unprefixed modifier. Just like 'privative', though, the feature 'approximative' is again too imprecise to distinguish these three prefixes' specific semantic contributions, and we therefore suggest an additional semantic distinction to differentiate them, that of 'disproximation'.

1.2 Approximation and 'disproximation': Contrastive examples and graphical representations

Our intuitions about subtle but noticeable semantic differences among the prefixes are fed by what we can observe in selected pairs of prefixes sharing a base, such as *quasi-philosophical* and *pseudo-philosophical*, or *near-biblical* and *pseudo-biblical*.[1]

(1) a. [The Japanese have] a **quasi-philosophical** devotion to the challenge of building a mechanized human. (COCA, 2006, MAG) (≈ 'an almost philosophical devotion')

   b. He opens this satirical thriller set in "our age" with a **pseudo-philosophical** commentary that includes this ludicrous statement: "In his understanding that love was not enough, in his acceptance of the necessity of the sacrifice of his own life to enable the future of those around him, Jesus is history's first, but not last, example of a suicide bomber." (COCA, 2007, NEWS) (≈ 'a commentary which is meant to come across as philosophical but that one shouldn't confuse with real philosophy')

(2) a. Last year the gulf coast got buffeted by a **near-biblical** onslaught of evil weather. (COCA, 2006, MAG) (≈ 'an onslaught of evil weather that took on proportions close to those of a plague or flood from the Old Testament')

   b. In "The Gospel According to the Son," he takes on the story of Jesus' life and tells it in the first person. The result is a choppy and unexciting telling of a tale we already know well. He uses **pseudo-Biblical** language that makes the story and the character of Jesus seem stilted. (COCA, 1997, NEWS) (≈ 'language that is meant to sound as though it belongs to (a hitherto lost part of) the Bible, but that in fact is not authentically biblical')

Such contrasts call for the need of introducing a 'dynamic' or 'orientational' perspective on their meanings. While *quasi-* and especially *near-* may express genuine approximation in the sense of 'coming close' to a standard or baseline level, *pseudo-* conveys the idea of being somehow fake and 'falling short' of a standard (cf. also Vassiliadou et al. 2023 and some further supporting observations from the literature that we will come to shortly).

---

[1] When we compare two differently prefixed adjectives involving the same base, it also becomes clear that the base can be understood in a somewhat different way depending on the prefix and the larger context. For instance, with *near-*, the adjective *biblical* typically means 'very great, severe', as in *a storm of near-biblical proportions*; with *pseudo-*, the same base is more likely to relate more closely to the Bible, its writing style, and/or its characters and the events narrated in it, as in *pseudo-biblical verses*, *pseudo-biblical declamations* or *a pseudo-Biblical virgin birth by Tattoine slave Schmi Skywalker*. For another example, *near-historic* usually means 'almost record-breaking', while *pseudo-historic* means 'not really historically accurate' or 'seemingly old and authentic, but actually of modern creation'. These facts are interesting and raise the question of whether the meaning of the prefix is responsible for picking out the relevant sense of the base. Of course, some prefixed combinations may be stored with a conventional meaning that no longer has to be 'computed' on the fly.

The Langackerian representations in Fig. 1 (adapted from Taylor's 2002: 220 representation of the adjective *tall*) aim to capture this distinction.
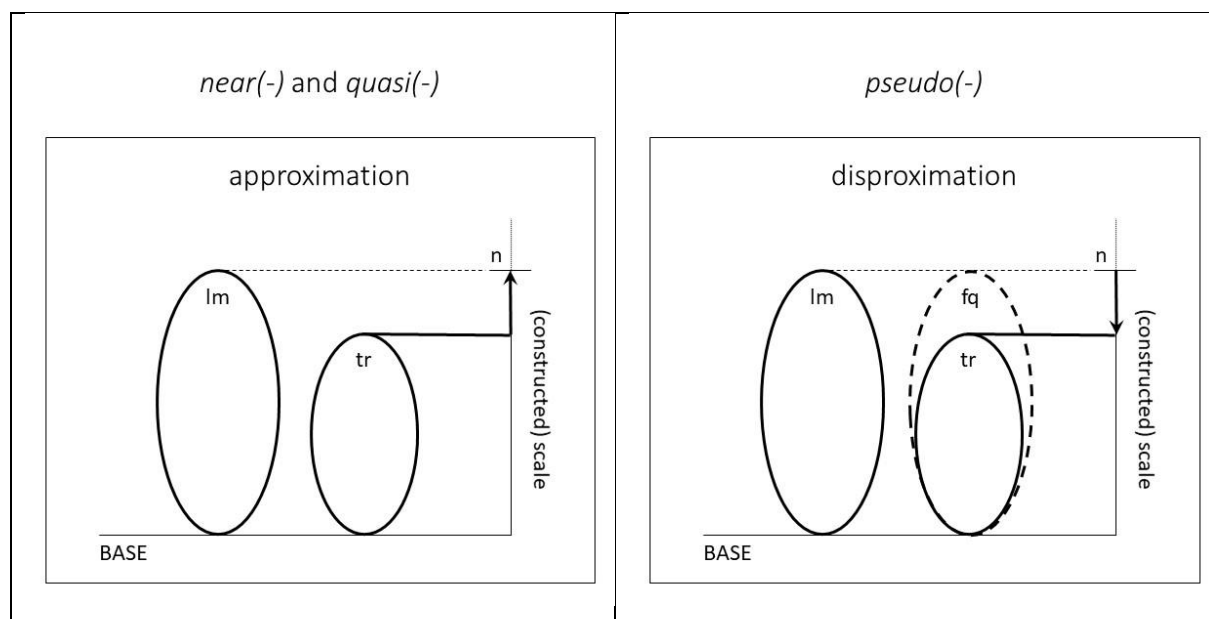
Fig. 1: *Near(-)* and *quasi(-)* profile the (true) approximation of a trajector (tr) to the norm (n) obtained by a landmark (lm), referred to by the morphological base, along some scale. *Pseudo(-)* profiles the 'disproximation' of a trajector, apart from its formal quale (fq), away from the norm obtained by a landmark.

In these representations, the 'BASE' at the bottom is not to be confused with the morphological base to which the prefix is added. The BASE is the context against which the relevant entity is conceptualized. It is left unspecified in these figures. (In the case of *near-complete*, the BASE could be 'quantity', for instance.) The scale along which the trajector moves, towards or away from a norm, is a scale which may yet have to be constructed in an ad hoc way (see Section 4.1). In the representation on the right, only the trajector's formal quale (fq) (i.e., what it looks like; cf. Pustejovsky 1995) obtains the norm set by the morphological base.

1.3 Previous accounts

For now, the distinction between *near-* and *quasi-* on the one hand and *pseudo-* on the other, shown in Fig. 1, is being presented as a hypothesis merely. There is some support for it, however, from descriptions in the literature. First, the definitions that the Oxford English Dictionary (OED) provides in the entries of these prefixes can be interpreted in light of the different orientations. Thus, for *quasi-*, we find (as part of the definition of its

use with an adjectival base) "almost, nearly, virtually", which is in line with the (true) approximation interpretation. For *pseudo-*, the OED lists a number of paraphrases that highlight the negative evaluation that comes with its non-scientific use: "Forming nouns and adjectives with the sense 'false, pretended, counterfeit, spurious, sham; apparently but not really, falsely or erroneously called or represented, falsely, spuriously'." The OED, surprisingly, does not include an entry for *near-* as a prefix, but in the entry for *near* as an adjective, it lists such compounds, written in two words or with a hyphen, as *near certainty*, *near cider*, *near famine*, *near-fascist*, *near-illiterate*, *near-miracle*, *near monopoly* and some others, where *near(-)* is glossed as "*nearly* the same as, falling *somewhat* short of, being an *acceptable* substitute for (the thing specified)" (italics ours – B.C., R.D. & S.H.).[2] While the gloss of "falling ... short of" is one that we gave above for *dis*proximation, the entire description suggests more of a genuinely approximative semantics, as should be clear from the hedge *somewhat* and the other modifiers we have italicized in the quotation.

Referring also to the OED, Bauer, Lieber and Plag (2013: 414) note that *pseudo-*, except in its scientific use (e.g., *pseudoarthrosis*, *pseudobinary*, *pseudonym*), "indicates something sham or less than genuine". The "less than" part is again interesting, as it suggests the negative idea of being deficient in some respect (hence, disproximation). These authors explicitly compare *quasi-* and *pseudo-* – though, unfortunately, they contrast neither with *near-*[3] – and, drawing on examples from COCA (Davies 2008–), they make this observation:

> A comparison of *quasi-* and *pseudo-* reveals that the two prefixes share the meaning component that the derivative does not refer to a genuine exemplar of its head's category. However, *quasi-* lacks the element of falseness that *pseudo-* generally carries. (Bauer, Lieber & Plag 2013: 414)

Dixon (2014: 170–171), too, contrasts *quasi-* and *pseudo-* only, remarking that they "are superficially similar but in fact exhibit a significant meaning difference" (p. 170). Not

---

[2] Another sense given by the OED for *near(-)* in compounds is "artificial", which we presume applies to combinations such as *near fur* or *near-silk*, referring to materials that resemble *sealskin* and *silk*, respectively.

[3] This may be due to the fact, noted above, that the OED does not explicitly recognize *near(-)* as a prefix. Bauer, Lieber & Plag (2013), do, however, mention "other potential rivals", namely the suffixes *-ish*, *-esque* and *-oid*, which like both *quasi-* and *pseudo-* form words that "denote something that is similar, but not identical, in shape or quality to what the base denotes" (p. 416). These suffixes are more like *quasi-* than like *pseudo-*, Bauer, Lieber and Plag (2013) still observe, in that they do not suggest falseness, but these authors provide no further insights on how to distinguish *quasi-* and these three suffixes among themselves, other than suggesting that their distribution is "perhaps a matter of register or domain". While these suffixes may not have an element of deceit, they certainly can be negatively tinged (cf. Sánchez Fajardo 2022).

having shed the etymological sources of these prefixes, *quasi*-X can be characterized as "having some characteristics of X but not being a full X" and *pseudo*-X, at least in its common, colloquial usage, as "pretending to be like X, or being similar to X, but in neither case being X" (Dixon 2014: 171). These meanings are illustrated, perhaps more humorously than accurately, by means of a couple of well-chosen contrastive pairs. For instance, a *quasi-cripple* is "someone who has some small thing wrong with them (say, missing two fingers from one hand) but not really so serious to justify the label 'cripple'" (Dixon 2014: 171) – note also the privative aspect in this 'definition' – while a *pseudo-cripple* is "someone who has nothing at all wrong with them but pretends to be a cripple (perhaps, so that they can take part in the para-Olympics)" (ibid).

In all these treatments, the element of falseness and counterfeiting only appears with *pseudo-*. It is this semantic component that likely gives rise to the idea that *pseudo*-X is disproximative, in our terms. This can be explained as follows. Given that the formal quale of an entity is what can be observed immediately (being the outer appearance of it), this is the starting point of any assessment of that entity. Once one sees through the pretence, the entity's real qualities are accessible and these will then be seen as clearly less impressive than what they were first made out to be. With *quasi*-X, we hypothesize, there is no misleading 'first appearance'. What we see is exactly what we get, but even though the entity does not reach a given norm (namely the norm needed to fully qualify as X), it does come close to it.

1.4 Questions, approach, scope

Even with a more refined semantic terminology, we believe that, to arrive at a suitable semantic characterization of the three prefixes under study (*near-*, *pseudo-* and *quasi-*), we need to let the data speak for themselves in the first place: the meanings of linguistic items should become apparent from their use. Specifically, our paper attempts to answer the following questions: Which semantic similarities and differences do we find among the three prefixes with respect to the bases they select, and does this allow us to come to a semantic characterization of the prefixes themselves? In other words, the main aim of this article is to find out what the meaning is of each of the three prefixes by looking at the types of bases each of them combines with.

The three prefixes of interest here are not meant to constitute an exhaustive set of privative and ap-/disproximative prefixes. Indeed, *half-* and *semi-* could also have been considered, in view of their privative and approximative meaning in, for instance, *half-dead*, *half-naked*, *semidark* and *semiprofessional*, all of which allow paraphrases with 'not X but almost X' (cf. Micheli 2023 on *semi-* in Italian). A Latinate alternative to *half-* is *demi-*, as in *demi-permanent*, as said of a non-permanent hair dye, one that, according to the definitions we found, lasts for up to 30 washes, while *semi-permanent* hair dye only lasts for about 5 washings; this suggests that *semi-*, at least in this particular combination, carries a meaning that is far from truly approximative. Another privative prefix is *fake-*, as in *fake-calm*, *fake-cheerful*, *fake-friendly*, *fake-innocent*, etc., but its meaning certainly is not obviously approximative, in that a paraphrase with 'almost' is not possible; rather, its meaning could probably be described as 'disproximative', in the sense explained above. Thus, if someone is said to act in a *fake-calm* way, the speaker intends to say that their behaviour only *looks* calm but is *different from* being truly calm. The prefix *sub-*, finally, often also has a disproximative meaning, as in *suboptimal* or *subhuman.*

We will now explain how we set out to find out what the three selected prefixes mean, based on the bases they appear with. Our approach is distributional-semantic. That is, when we consider the bases of the prefixes, these are clustered according to their shared contexts of use, not according to their similar morphological properties or language origins. This approach means that some potentially relevant questions are left out of consideration. For instance, both *quasi-* and *pseudo-* are non-native (going back to Latin and to Greek, respectively), in contrast to Germanic *near-*, and, as a reviewer suggested, it would therefore be interesting to find out whether these different origins are reflected by different shares of native and non-native bases and, if so, whether this might also help us in characterizing the differences among these prefixes. We leave this for future research (but see Section 4.4 for some tentative observations) and happily invite other researchers to explore this.

2. Data and methods

2.1 Corpus data and annotation scheme

We retrieved the data from a CQP-transformed version of the commercially available *Corpus of Contemporary American English* (COCA; Davies 2008–; 450 million words, AmE, 1990–2012). Multiple queries were run, in order to include the different bases with which each prefix potentially combines (i.e., adjectives, nouns, verbs, and adverbs) and to account for spelling variation (i.e., one word, hyphenated, or two words). No restrictions on genre or time period were imposed beforehand. After retrieval, the data were cleaned semi-automatically and the prefixed forms were annotated for word class, syntactic distribution (specifically, in case the prefixed word is an adjective, whether it is used attributively or predicatively), and spelling type (hyphenated, single word, two words).[4] Table 1 illustrates the annotation.

Tab. 1: Annotation used for the tokens extracted from COCA

Token	Word class	Syn. distr.	Explanation
GMACC, which has an operating budget of $297,000, is a **quasi-public** agency coordinating Georgia' s campaign to save its bases. (COCA, 2003, NEWS)	A	Attr	adjective, attributive
Again, these formulas have the appearance of being scientific because they use estimates of reliability, standard errors, and the like. However, they are **pseudoscientific** because the numbers do not relate to a valid construct of Learning Disabilities (LD). (COCA, 2006, ACAD)	A	Pred	adjective, predicative
The influence of age on the infants' performance was tested by distributing infants in the three age categories. The 36 infants were **pseudo-randomly** distributed into three experimental groups according to the type of reinforcement, with an equal number of boys	ADV		adverb

---

[4] Although an investigation of spelling was beyond the scope of the present paper, future research should establish whether there are any patterns to be observed in the presence or absence of a hyphen. For instance, does *pseudo* occur as a free morpheme more often when followed by a noun than when followed by an adjective (cp. *a pseudo philosopher* and *pseudo-philosophical*, to use examples found in WordNet)? Also, does *near* connect to the base with a hyphen relatively more often when used in an attributive adjective (e.g., *a near-perfect game*) and appear as a free morpheme relatively more often when the adjective is predicative (e.g., *The game was near perfect*)?

Token	Word class	Syn. distr.	Explanation
and girls and a similar age distribution in each group. (COCA, 2007, ACAD)			
Prosecutors pull detectives off important investigations just to serve subpoenas. These and other problems can be traced to the **near collapse** of the county sheriff's office and a tug of war between two of the county's most stubborn elected officials. (COCA, 1998, NEWS)	N		noun
[R]eligious divisions that had been similarly subdued beneath an official atheism have been revived and **pseudo-ethnicized** (thus, for the Serbs the Muslim Slavs are "Turks"). (COCA, 2000, ACAD)	V		verb
Une pareille lecture a vue des connexions imaginaires les plus subtiles, telles qu'elles peuvent se produire par example dans la poesie… m'inquiete quelquefois comme si elle relevait d'un domain **quasi-religieux** d'interdit. (COCA, 1990, ACAD)	FW		foreign word
One night a woman came into the bathroom and caught me hunched over like **Quasimodo**, staring intently at the drains , my hands full of dead moths. (COCA, 1990, FIC)	FALSE.POS		(other type of) false positive
He then moved to Paris and lived under the **pseudonym** Sebastian Melmoth. (COCA, 2003, FIC)	CF		combining form
Nor are they all-powerful but rather operate according to their roles in the celestial realm. A sixth-century writer we call **Pseudo-Dionysius** outlined nine choirs of angels and their job descriptions, which are cross-referenced with biblical examples of each. (COCA, 2009, MAG)	PN		proper noun
[A]nd, as a result of the, I guess it was a **near encounter**, the horse broke the harness and ran away, and when Black came back to his blacksmith shop, where he had built the car, around the corner […]. (COCA, 1996, SPOK)	?		ambiguous/unclear (either semantically or morphosyntactically)

Foreign words (e.g., *quasi-religieux*, *pseudo-identitaire*) as well as other false positives (e.g., *Quasimodo*) were discarded. We also excluded proper nouns (e.g., *Pseudo-Dionysius*), clearly technical terms (e.g., *pseudo R^2*, *pseudoephedrine*), and combining forms (e.g.,

pseudonym) from the investigation. Even though these still carry traces of the approximative (or disproximative) meaning of the respective prefixation patterns, these items may not reflect the way the prefixes discussed here are used as productive elements in more spontaneous registers. Bauer, Lieber and Plag 2013 (414–415) similarly distinguish the "colloquial usage" and the "scientific usage" of *pseudo-*, remarking that the latter "lacks the evaluative attitude". That said, it is hard to make a principled distinction between non-technical and technical terms. For one, a single lexical item, used in the same subcorpus, may either exhibit the subjective, evaluative use of pejoration (e.g., 3a) or belong to the purely scientific level, conveying no depreciation at all (e.g., 3b):

(3) a. Indeed, very few successful socialist propagandists ever bothered to focus on the empirical case for socialism. Rather, when trying to sell socialism as a policy or a movement, its preachers testify about "social justice," "humane policies," "fairness," and "equality." In short, socialism—be it Marxist, Fabian, nationalistic, progressive—is merely one of many **pseudo-empirical** rationalizations of the deeper psychological impulse of Blair's "social-ism." The true case for socialism is not to be found in GDP or employment numbers, but in the promise of leaping out of History into a better society where we are all loved and respected as members of the same family. (COCA, 2010, ACAD)

b. You, Y., and Rao, J.N.K. (2002). A **pseudo-empirical** best linear unbiased prediction approach to small area estimation using survey weights. Canadian Journal of Statistics, 30, 431–439. (COCA, 2019, ACAD)

This implies we cannot simply assume that in academic discourse, *pseudo-*, when combined with a base of Latinate or Greek origin, will be technical. For another, it is also too simplistic that even in non-technical discourse, *pseudo-* always comes with a pejorative interpretation. Consider this example:

(4) Last year's Gear S2 had a sleek, **pseudo-futuristic** vibe – so much so that the white model I reviewed looked like a prop straight out of THX 1138. Samsung ditched that clean aesthetic this time around – the S3 Frontier rocks a rugged look, with a knurled, rotating bezel and a chunky stainless-steel body. (COCA, 2016, MAG)

The uses of *pseudo-* thus seem to form a cline from colloquial and negatively-evaluative (e.g., ***Pseudo-artsy*** *film featuring dribble called from a dark chat room*), to neither very colloquial nor scientific, and apparently (quite) neutral (e.g., *They went past a red,* ***pseudo-Gothic*** *church, past a small bridge and railway tracks*), to scientific and non-evaluative (e.g., *The test signal is generated from a digital* ***pseudo-random*** *noise (PN) generator*)

(examples again taken from COCA). In view of these observations, we may decide to leave the latter in after all in subsequent analyses.

Finally, we discarded ambiguous or unclear cases. For instance, for *near encounter*, we could not unambiguously establish whether *near* expresses spatial proximity ('close'; 'close-by'), as in *near miss* (more examples will be given shortly), or approximation ('almost'), as in *near collision*; in some other cases (e.g., *pseudo-conservative*), it was not clear from the context of the prefixed word whether the base was a noun or an adjective.

The extracted examples with *near(-)* were especially prone to noise. The number of false positives reached about 5,890 cases out of the 11,614 examples retrieved; that is, just over half of the cases were false positives. The majority of those (about 4,500 tokens) involved *near(-)* meaning 'proximal', 'in proximity to', 'in the vicinity of' or 'close(st) to (a reference point)'. In these cases, *near(-)* is the opposite of 'far (from)' and does not carry the meaning of 'almost' or '(metaphorically) approaching'. Examples are *near(-)* in *the near future*, *the near side of moon*, *Near Eastern*, *near-surface currents*, or *near-sighted*. Also excluded were examples in which *near* occurs as a literally used preposition (e.g., *near home*, *near schools*, *near Silicon Valley's tech companies*) or a metaphorically used one (e.g., (*to be*) *near bankruptcy*, (*to be*) *near retirement*, (*to be*) *near tears*). A third category of excluded examples involved those with *near* meaning 'close', 'with hardly any margin or excess quantity (e.g., of time, of space)', 'with little difference (between competing parties, etc.)', rather than 'almost' (e.g., *a near escape*, *a near tie*, *a near-run thing*, and *a near miss*, which does not mean 'almost a miss', but 'a collision or a shot which comes close to hitting (target)', 'a failure, but not by much').

Something that presented quite an annotation challenge was when the base was not clearly a noun, an adjective, or another part of speech. For instance, the underlined base in *near record snowfall* and *near-textbook recipe for disaster* is a noun that appears to be used as an adjective; in that case, 'A or N' was given as annotation code (rather than '?'). Likewise, in partly adjectival and partly verbal cases such as *near-blinding*, *near-scalding* and *near-endangered*, we used the code 'A or V'. Cases that could be considered partly nominal and partly verbal, such as *near-doubling*, *near-tripling*, *near-drowning* and *near vanishing*, were given the code 'N or V (gerund)'. Finally, still for *near(-)* at least, we used a special annotation code when the prefix was used with an adjective or noun that formed

part of a nominal compound, as in *near civil war* and *near-capital offense* (annotation code: 'A of AN compound') and in *near plane crash* and *near-death experience* ('N of NN compound'). We will return to such compounds in Section 4.2. Table 2 provides an overview of the raw frequencies for each prefixation pattern with their different bases.

Tab. 2: Frequency distribution of *pseudo-*, *quasi-*, and *near-* with different bases and spelling variants; COCA

Morphological base	Variant	Prefix			Total
		pseudo(-)	*quasi(-)*	*near(-)*	
adjectival	x A	12	40	397	
	xA	179	117	54	
	x-A	311	1,224	1,984	
	Total	502	1,381	2,435	4,318
nominal	x N	115	68	1,201	
	xN	448	106	32	
	x-N	601	452	1,022	
	Total	1,164	626	2,255	4,045
adverbial	x ADV	0	0	1	
	xADV	3	5	0	
	x-ADV	3	19	0	
	Total	6	24	1	31
verbal	x V	0	0	0	
	xV	0	0	0	
	x-V	1	1	5	
	Total	1	1	5	7
	residue	1,411	201	6,918	8,530
	TOTAL	3,084	2,233	11,614	**16,931**

The category 'residue' contains compounds, foreign words, combining forms, false positives, and ambiguous cases. Fig. 2 presents the results presented in Table 1 more clearly.
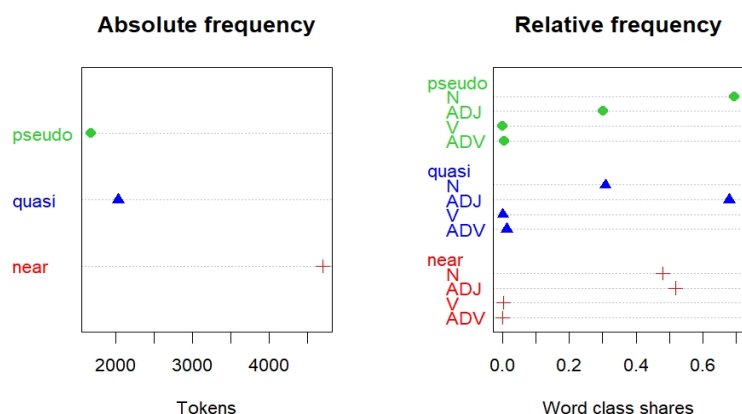
Fig. 2: Absolute frequency of words prefixed by *pseudo-*, *quasi-* and *near-* after clean-up (left) and relative frequency of prefixed words by the syntactic category of their base (right)

Note that *near-* is by far the most frequently used prefix of this triplet. It can also be observed that about 70% of words with *pseudo-* are nouns and only about 30% of such words have an adjectival base, while the opposite pattern holds for words with *quasi-*. So, although both *pseudo-* and *quasi-* are learned prefixes (having a Greek and Latin origin, respectively) and have somewhat similar meanings – indeed, they are sometimes seen as synonyms – their morphological behaviour is rather different. With *near-*, the distribution of nominal and adjectival bases is roughly fifty-fifty. All three prefixes hardly combine with verbs or adverbs. The subsequent analyses are therefore limited to nominal and adjectival bases.

2.2   Semantic vector spaces

For investigating the semantic characteristics of the patterns under investigation in more detail, we use distributional semantics, an explorative method that has become ever more popular in recent years (see, e.g., Levshina & Heylen 2014; Perek 2016, 2018; Hilpert & Perek 2016). Distributional semantics draws on the idea that words that are similar in meaning will occur in similar contexts – a hypothesis that has been confirmed in a number of distributional-semantic studies. Thus, similarities and dissimilarities between linguistic items can be characterized with the help of their collocates. In a 'bag-of-words' approach, the collocates of each word are used as so-called semantic vectors, and the similarities between them are quantified using an association measure such as Mutual Information and a measure of similarity such as Cosine similarity. In our approach, we follow Levshina's (2015) operationalization of the method, using Positive Pointwise Mutual Information (PPMI) as association measure and Cosine similarity as measure of similarity.

    As we are interested in the semantics of the head constituents, we queried COCA for the adjectives and nouns that are attested in combination with *near-*, *quasi-* and *pseudo-*, and the concordances obtained in this way were used to extract the collocates of each lexeme in a window of five words to the left and to the right.[5] So-called stop words, i.e., words that are not very informative for our purposes due to their very high frequency, were excluded from the collocates using the R package *stopwords* by Benoit et al. (2021). We use multi-dimensional scaling (MDS; see e.g., Wheeler 2005) in order to be able to represent

---

[5] In a similar vein, for Dutch *fake*-type morphemes, Van Goethem and Norde (2020) carried out a semantic vector-space analysis on the element to the immediate right of these prefixes.

the results in two-dimensional space. To keep the plots readable, only items attested at least 5 times in the entire dataset are taken into account.

3. Results

3.1 Number of shared bases

Table 3 below shows for each pair of prefixes the number of morphological bases shared.

Tab. 3: Number of bases shared by pairs of prefixes, per word class of the base

pair	adjectival base	nominal base
near- and *quasi-*	80	36
near- and *pseudo-*	35	28
pseudo- and *quasi-*	84	47

As can be seen by comparing the first and second lines of Table 3, *near-* shares more bases (both adjectival and nominal ones) with *quasi-* than it does with *pseudo-*. This seems to confirm that *near-* and *quasi-* are more closely related to each other than *near-* and *pseudo-*. But then again, *quasi-* and *pseudo-* turn out to share even more bases among themselves, calling into question the assumption that *near-* and *quasi-* pattern alike and are both different from *pseudo-*. We suggest that the number of shared bases, while to some extent indicative of the semantic relatedness of the prefixes, can only tell part of the story. A more in-depth exploration of the sorts of bases the prefixes attach to is called for.

3.2 Semantic vector-space results

Figures 2a and 2b show the results of a semantic vector-space analysis for the lemmas occurring in the three constructions under investigation (*near*-X, *pseudo*-X, *quasi*-X). Figure 2a shows the results for adjectives, Figure 2b for nouns.[6] The results allow for (tentatively) identifying some semantic clusters of nouns and adjectives that combine with the three affixes. They also suggest that while there is much overlap between the semantic domains, there are a few niches that are almost exclusively occupied by *quasi-* (in blue) and especially *near-* (in red).

---

[6] An interactive version of these plots can be found at https://hartmast.shinyapps.io/ApproximativeVector-Spaces (accessed 1 July 2022). Note that some adjectives also occur in the nominalized form (e.g., *the pseudoscientific* as a person collective), hence some lexemes occur in both plots.

On the left-hand side of Figure 2a, we see many lexemes that have to do with institutions (*governmental*), society (*democratic*), as well as ideologies and/or schools of thought (*socialist*, *marxist*). On the right-hand side, we find more concrete state adjectives (*black, white, dead*). The middle ground and especially lower half of Figure 2a are occupied by a fairly heterogeneous set of adjectives, many of which are scientific terms (*experimental*, *isotropic*, *academic*), while others are more to do with the 'esoteric' (*mythic*, *magical*). Especially the lexemes in the middle area of the plot often carry a certain positive or negative semantic prosody (*stellar*, *hysterical*).

The nouns that combine with *pseudo-* and *quasi-* (Figure 2b) show an even clearer connection to the academic domain, and to a certain extent, this also goes for *near-*. This might indicate that the three word-formation patterns are characteristic of specific (academic) text types or at least specific 'learned' lexical items. Apart from this, we find similar semantic groups in Figure 2b as in Figure 2a, e.g., social/institutional terms or terms referring to ideologies (*government, democracy, capitalism*) but also substance (*beer, liquid*) and state nouns (*death*).
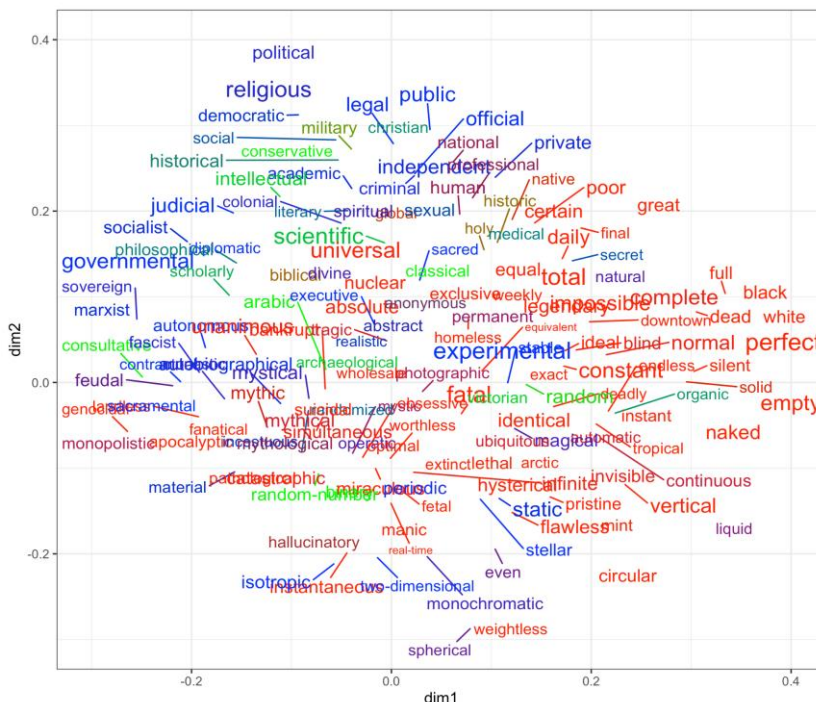
Fig. 2a: Results of a semantic vector-space analysis for the adjectives attested in combination with *near-*, *pseudo-*, and *quasi-*. The colours indicate the prefixes each item preferentially occurs with: red stands for *near-*, green for *pseudo-*, blue for *quasi-*, with different shades in-between representing 'promiscuous' items that occur in more than one of the three constructions.
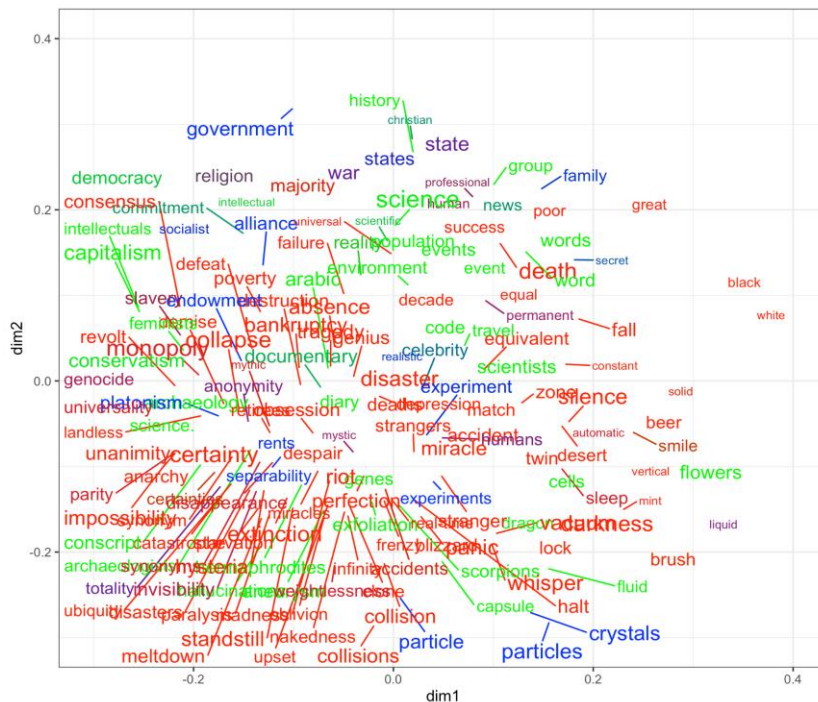
Fig. 2b: Results of a semantic vector-space analysis for the nouns attested in combination with *near-*, *pseudo-*, and *quasi-*. They are represented in the same way as in Figure 2a.

4. Discussion

Apart from what we can gain from the number of shared bases and the semantic vector-space analyses, the stage of preparing the data for analysis (see again Section 2.1) already allowed us to make some mostly qualitative observations. Some of these assorted findings are discussed in Sections 4.1 to 4.3. We will then return to the findings about shared bases in Section 4.4 and to the semantic vector-space analyses in Sections 4.5 and 4.6. A close reading of some selected examples with *religious* as a base, shared by the three prefixes, will be provided in Section 4.7.

4.1 Constructionally induced scalarity

With *near(-)*, a scalar interpretation may be constructionally induced, even when scalarity is not clearly part of the meaning of the noun:

(5)   a.   he is a near skeleton

      b.   the beans had dissolved into a near puree

      c.   my breathing had quickened to a near pant

What we find here is a sort of coerced reading. *Near*-X is so frequently used with a scalar interpretation, especially but not exclusively with adjectival bases (e.g., *near-complete*, *near-blind*, *near-perfection*), that nouns whose meaning is not standardly seen as forming a point high up on a cline (*skeleton*, *puree*, *pant* in the examples above) can nonetheless easily be perceived as doing so (*skeleton* indicates a high point on a scale of thinness, *puree* on a scale of mushiness, *pant* on a scale of frequency of in- and exhaling).

4.2 Compounds: Scope relations and infelicitous hyphenation

Near(-) often occurs in compounds, with different scope possibilities. Compare (6) and (7), where the added brackets indicate what belongs together:

(6)   a near [ghost town], a near [heart attack], near [dirt roads], a near [car accident]

(7)   a [near-death] experience, a [[near zero]-emission] plant

The bracketing structure of these examples is not always clear, though: should it be *near-*[*starvation wages*] or [*near-starvation*] *wages*? In some cases, hyphenation may provide a pointer as to how the pattern should be understood, but the following examples show that hyphenation is not always 'logical' (square brackets again added):

(8)   near-[photo finish], near-[rubber stamp approval], near-[light speed], near-[folk hero], near-[world record]

What makes hyphenation clumsy in these cases is that *near* is typographically represented as forming a unit with the next element only (e.g., *near-rubber*) while it forms a unit with the whole compound (e.g., 'what almost amounts to rubber stamp approval'). This suggests that there are more factors influencing the use of hyphens, as opposed to whitespace or univerbation (see Sanchez-Stockhammer 2018), which is why hyphenation alone cannot be used for clearly disambiguating between scope alternatives in such cases.

4.3 Tiny niches of productivity: An example

While annotating the data, it became clear that prefixes can be used productively with similar bases in extremely low-level patterns. One example is *near(-)* in combination with nouns referring to a manner of speaking. Thus, apart from *near whisper*, which appears with sufficient frequency to be visible in the plot, we also found items which themselves have too low frequency for this – usually, they are hapaxes – but which together do form a

discernable pattern: *near growl, near-holler, near monotone, near murmur, near pant, near shout, near-sneer, near snort, near-wail.* In an example such as *"Do not argue with me," she said in a near monotone*, we arguably see the following conventionalized pattern at work:

(9)   "[Quotation]" X said in/with a near N$_{\text{manner of speaking}}$.

4.4 Shared bases in numbers

We noted in Section 3.1 that *near-* and *quasi-* share a larger number of bases among themselves than the former does with *pseudo-*. This finding seems to confirm our hypothesis that *near-* and *quasi-* share a semantic feature, presumably then that of approximation, while *near-* and *pseudo-* are semantically different, which could then be taken to be due to *pseudo-* expressing disproximation. Our hypothesis, however, only finds partial support from the data, because *quasi-* and *pseudo-* share even more bases than *near-* and *quasi-* do. At this stage, we cannot fully account for these results. Here are nonetheless some tentative considerations.

First, *near-* is of Germanic origin while both *quasi-* and *pseudo-*, as noted before, are learned prefixes. This could in part explain why *quasi-* and *pseudo-* share so many bases: these are typically Latinate or Greek-derived items one could expect to find in more formal varieties of English: *democratic, futuristic, intellectual, rational, theological*, etc. Bauer, Lieber and Plag (2013: 416) note that both *quasi-* and *pseudo-* are "eligible for the coinage of a new scientific term, and we do find both, perhaps with a preference for *pseudo-*." Compared to these two prefixes, *near-* more often appears with bases of Germanic origin, such as *dead, naked, flawless*, and so on. That *near-* and *quasi-* nonetheless appear to share many bases (e.g., *annual, free, magical, nude*, etc.) then possibly does speak to their similarity.

Second, if a pair of prefixes share a large number of bases, this need not be taken to mean that these prefixes are *identical* in meaning. This adds a caveat to the preceding point about the relatedness of *near-* and *quasi-*. Likewise, even if *quasi-* and *pseudo-* have many bases in common, this in itself does not suggest that they are close synonyms. After all, they could in principle express related but rather different (indeed opposite) meanings with respect to that base, as is the case for pairs like *sub-* and *supra-* (cf. e.g., Gries & Otani 2010 on the behavioural profiles of antonyms). We do not have sufficient evidence to claim, however, that this is actually also the case for *quasi-* and *pseudo-* (but see Section 4.7 for at

least one convincing illustrative case), although it is certainly something that needs to be considered as a possibility.

Third, in this respect, if *near-* and *pseudo-* were each other's antonyms (expressing approximation vs. disproximation), then we would expect them to share many bases. That this is not the case may not only be due to the fact that these prefixes have their origin in different language families (cf. our first point noted above) but could also be explained by the fact that *near-* and *pseudo-* are not perfectly opposite in meaning. In ordinary (non-technical) uses of *pseudo-*, there is an evaluative semantic component relating to fakeness, dishonesty, pretence, etc., for which there is no clear positive counterpart in the use of *near-* (cf. again Section 1). Thus, we can describe something as *pseudo-scientific* (i.e., 'only pretending to be scientific') while it is not clear what it would mean for something to be *near-scientific*. It is possible, however, to claim something with *near-scientific certainty* or to carry out an action with *near-scientific precision*, but then the whole adjective-noun combination (*scientific certainty*, *scientific precision*) is felt to be modified semantically by *near-* (cf. also Section 4.2 for spelling-scope mismatches).

4.5 Kinds of bases: Further observations from the vector-space analyses

The semantic vector-space analyses hint at a few distinct clusters in which each of the prefixes is used. Selecting adjectival bases and spoken register, for instance, we can see that *near-* selects bases with negative prosody (e.g., *near-deadly*, *near-fatal*, *near-suicidal*, *near-apocalyptic*), although we also find it with bases that are neutral or positive (e.g., *near-invisible*, *near-universal*, *near-perfect*). This prefix is clearly the most productive of the three so far investigated. Using the same settings, *pseudo-* only appears with sufficient frequency and without much competition from other prefixes in the well-known combinations *pseudo-intellectual* and *pseudo-scientific*, though it also shares bases like *religious* and *medical* with *quasi-*. Still with the same settings, *quasi-* occurs with a larger range of bases than *pseudo-* to form such adjectives as *quasi-diplomatic*, *quasi-governmental*, *quasi-judicial*, *quasi-public*, *quasi-private*, all of which have a precise, legal-administrative definition. These are, in a sense, approximative but not evaluative. Yet, we also find combinations such as *quasi-religious*, whose base is shared with both *pseudo-* and *near-*, and *quasi-*

unanimous, *quasi-universal*, *quasi-divine*, *quasi-mystic*, *quasi-complete* and *quasi-normal*, all of which also appear in the *near-* variant, and more frequently so.

4.6 Approximation and disproximation: Visible in the data?

The semantic vector-space analyses, as well as an explorative look at the distribution of the bases between the different prefixes, provide *some* support for the hypothesis that *near-* and *quasi-* are more similar to each other than either of these is to *pseudo-*, in terms of the 'dynamic' or 'orientational' interpretation. We find that *near-* shares many more of its bases with *quasi-* than it does with *pseudo-*, which is in line with our intuition that *quasi-* and *near-* are semantically similar, while the meaning of *pseudo-* is apparently different. However, this in itself does not confirm beyond reasonable doubt that the former two express approximation and the latter disproximation. If *pseudo-* is disproximative in meaning, involving a sense of 'pretending to be without actually being' (cf. Section 1), we can predict that this prefix does not occur with bases that have negative prosody. This appears to be the case: we do not encounter combinations such as ?*pseudo-fatal* or ?*pseudo-catastrophe* in our dataset.

Puzzlingly, however, we also do not find combinations such as ?*pseudo-perfect* (unless in a technical sense: *pseudoperfect numbers*) and we do not find many negative-prosody bases with *quasi-* either (ˀ*quasi-fatal*, ˀ*quasi-catastrophe*).[7] The latter observation suggests that in English, *near-* is the default prefix for approximative meaning and that the use of *quasi-* is, by comparison, quite restricted, unlike for instance *quasi-* in French, which even has an approximative adverb variant (*quasiment* 'almost').

---

[7] In larger corpora, such as enTenTen20 (over 36 billion words in size), available via SketchEngine, some of the combinations to which we assign a question mark may be found, as noted by a reviewer, who cited the ones below:

(i)   … the **pseudo-perfect** setting from his childhood dreams. (adult-fanfiction.org)
(ii)   … missing the crown of thorns around his heart and the **quasi-fatal** wound … (egodeath.com)

Though rare, these examples are in line with our intuitions. *Pseudo-perfect* hints at a deceptive and/or unrealistic perception of perfection. *Quasi-fatal* means 'as good as fatal', 'practically fatal'. On enTenTen20, still no example was found of *pseudo-fatal*. This word is not necessarily unacceptable but can only be used in special situations, where it makes sense for something to look fatal without actually being so, as in this rare web-attested example referring to the well-known Milgram experiments in psychology:

(iii) With the confidence and supervision of these "Scientists/Doctors" these people were administering **pseudo-fatal** doses of [electricity] just because someone in a lab coat said it was the right thing to do. (https://www.physicianassistantforum.com/topic/11888-what-to-do-if-md-is-wrong, accessed 31 January 2023)

4.7 *Near-/quasi-/pseudo-religious*: Close reading

We end our paper with a close reading of, to our mind, representative examples of the three prefixes attached to the same base, allowing us to elucidate the differences between each prefix:

(10)   a.   As the argument progresses, his rhetoric rises to a **near-religious** fervor that is hardly orthodox for literary criticism. (COCA, 2010, NEWS)

       b.   [...] the self-identified "third-wave Confucians" have betrayed Kongzi's Way in their eagerness to embrace first Buddhist and later Western tenets, while elevating Western science and a range of postmodern theories to **quasi-religious** status. (COCA, 2016, ACAD)

       c.   You take a vulnerable kid, an addict, drop him in a mess of death and doubt, it won't be long before he starts looking for something to cling to. In comes Mama Celia with her authentic Mexican cuisine, side of **quasi-religious** voodoo bullshit and voila. You bet your ass she's got her hooks in him. (COCA, 2016, TV)

       d.   [...] Mitchell, who sang through his court proceedings and spouted **pseudo-religious** gibberish, was a skilled con man who was largely faking his delusions. (COCA, 2011, NEWS)

In (10a), where *religious* combines with *near-*, the context item *rises to* is in line with the notion of approximation. Of the 27 tokens in COCA containing *near-religious*, 6 are followed by the (otherwise low-frequency) noun *fervor*, which is also high on a scale of emotions. Some of the other nouns following *near-religious* are *experience* (6 tokens), *devotion* (2 tokens), and *ecstasy* (2 tokens). The author of (10a) seems to characterize someone's rhetoric as something that *comes close* to religious fervor. In both (10b) and (10c), we find the adjective *quasi-religious*. In (10b), its use is clearly approximative – note the use of the verb *elevate*, again suggesting movement to a particular standard. In (10c), *quasi-religious* is used in a rather different way, but this use is perhaps not necessarily *dis*proximative. Its meaning could be, as in (10a), 'a sort of', 'a kind of' and *quasi-* in this example thus comes close to functioning as a hedge. *Bullshit* here means something like 'stupid talk that I do not like' – the speaker seems to struggle to find the appropriate term. In other words, the meaning of *quasi-religious voodoo bullshit* could be 'stupid voodoo talk that is almost/kind of religious in nature'. Note that the context contains *authentic* (albeit applied to another noun), so the meaning here is not necessarily 'voodoo bullshit that falls short of being religious'. That said, *quasi-* actually *can* have a disproximative interpretation, as in (11):

(11)   Available wherever dubious, **quasi-scientific** self-help books are sold. (COCA, 2009, TV)

Here, the adjective *dubious* makes it clear that the language user does not find that self-help books are of an almost scientific quality; rather, the interpretation is that they fail to meet the standard of proper scientific rigour. Finally, in (10d) above, the claim that *pseudo-* has a clearly disproximative interpretation is supported by context items such as *spouted*, *gibberish*, *con man*, and *faking*. A close paraphrase of *pseudo-religious gibberish* is not 'gibberish that is almost/kind of religious in nature'; instead, the meaning in context here is more likely 'fake-religious statements which make no sense, and which we can therefore call gibberish'. Other nouns that may be found to follow *pseudo-*ADJ (e.g., apart from *gibberish*, we find again *bullshit*, but also *claptrap*, *drivel*, *garbage*, *gobbledygook*, *jargon*, *nonsense* and *pap*, none of which occur in COCA after *quasi-*ADJ) are revealing in that they provide strong cues for the negative connotation typically associated with *pseudo-*.

5. Conclusion

We have come some way toward differentiating the use of the English privative prefixes *near-*, *pseudo-* and *quasi-*. All three prefixes occur with adjectival and nominal bases. By looking at the bases, we have seen that *near-* and *quasi-* share more bases with each other than the former does with *pseudo-*, lending partial support to the assumption that the *near-* and *quasi-* are semantically similar to each other and that *pseudo-* is different. We have also seen, however, that *quasi-* and *pseudo-* share many bases (even more so than *near-* and *quasi-* do among themselves), although this could be an effect, in part, of both these affixes' non-Germanic origin. We have also found that *near-* is a default, highly productive prefix. *Pseudo-*, at least in nontechnical discourse, is least productive. *Quasi-* displays a little more productivity than *pseudo-*, appearing for instance in combinations to do with political/ideological leanings.

Work that lies ahead will have to be of a both quantitative and qualitative nature. For instance, we should conduct a semantic vector-space analysis not just on the bases but also on the prefixed words as a whole, so as to get more insight into the ways in which these words are used. In an attempt at proving that *pseudo-* is disproximative (unlike *quasi-*, or at least *more* so than *quasi-* and definitely unlike *near-*), we could then zoom in on the kinds of nouns that we find after an adjective prefixed in a particular way. Such detailed

analyses might then include the observation, for example, that *pseudo*-ADJ may be followed by nouns such as *drivel* or *nonsense*, as we anecdotally observed here, and possibly by other context items that are suggestive of a depreciative semantics.

References

Bauer, Laurie, Rochelle Lieber & Ingo Plag. 2013. *The Oxford reference guide to English morphology*. New York: Oxford University Press.

Benoit, Kenneth, David Muhr & Kohei Watanabe. 2021. *Stopwords: Multilingual Stopword Lists*. Manual. https://CRAN.R-project.org/package=stopwords.

Boleda, Gemma, Eva Maria Vecchi, Miquel Cornudella & Louise McNally. 2012. First order vs. higher order modification in distributional semantics. *Proceedings of EMNLP*, Jeju Island, Korea, 1223–1233.

Cappelle, Bert, Pascal Denis & Mikaela Keller. 2018. Facing the facts of *fake*: A distributional semantics and corpus annotation approach. *Yearbook of the GCLA* 6. 9–42.

Davies, Mark. 2008. *The Corpus of Historical American English*. https://www.english-corpora.org/coca/.

Dixon, Robert M. W. 2014. *Making New Words: Morphological Derivation in English*. Oxford: Oxford University Press.

Gries, Stefan Th. & Naoki Otani. 2010. Behavioral profiles: A corpus-based perspective on synonymy and antonymy, *ICAME Journal* 34. 121–150.

Hilpert, Martin & Florent Perek. 2015. Meaning change in a petri dish: Constructions, semantic vector spaces, and motion charts. *Linguistics Vanguard* 1(1). 339–350.

Kamp, Hans. 1975. Two theories about adjectives. In Edward L. Keenan (ed.), *Formal Semantics of Natural Language*, 123–155. Cambridge: Cambridge University Press.

Levshina, Natalia. 2015. *How to Do Linguistics with R. Data Exploration and Statistical Analysis*. Amsterdam & Philadelphia: Benjamins.

Levshina, Natalia & Kris Heylen. 2014. A radically data-driven Construction Grammar: Experiments with Dutch causative constructions. In Ronny Boogaart, Timothy Colleman & Gijsbert Rutten (eds.), *Extending the Scope of Construction Grammar*. Berlin & New York: De Gruyter.

Micheli, Silvia. 2023. The emergence of approximative values in Italian *para-* and *semi-*. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1). 156–176.

"near, adj. (and n.)". 2022. *OED Online*, Oxford University Press. www.oed.com/view/Entry/125544 (accessed 30 January 2023).

Nayak, Neha, Mark Kowarsky, Gabor Angeli & Christopher D. Manning. 2014. *A Dictionary of Nonsubsective Adjectives*. Stanford: Stanford University Reports.

Perek, Florent. 2016. Using distributional semantics to study syntactic productivity in diachrony: A case study. *Linguistics* 54(1). 149–188.

Perek, Florent. 2018. Recent change in the productivity and schematicity of the *way*-construction: A distributional semantic analysis. *Corpus Linguistics and Linguistic Theory* 14(1). 65–97.

"pseudo-, comb. form". 2022. *OED Online*, Oxford University Press. www.oed.com/view/Entry/153742 (accessed 30 January 2023).

Pustejovsky, James. 1995. *The Generative Lexicon*. Cambridge, Massachusetts: MIT Press.

"quasi-, comb. form". 2022. *OED Online*, Oxford University Press. www.oed.com/view/Entry/268293 (accessed 30 January 2023).

Sánchez Fajardo, José A. 2022. *Pejorative Suffixes and Combining Forms in English*. Amsterdam & Philadelphia: Benjamins.

Sanchez-Stockhammer, Christina. 2018. *English Compounds and their Spelling*. Cambridge: Cambridge University Press.

Taylor, John R. 2002. *Cognitive Grammar*. Oxford: Oxford University Press.

Van Goethem, Kristel & Muriel Norde. 2020. Extravagant "fake" morphemes in Dutch: Morphological productivity, semantic profiles and categorical flexibility. *Corpus Linguistics and Linguistic Theory* 16(3). 425–458.

Vassiliadou, Hélène, Francine Gerhard-Krait, Georgia Fotiadou & Marie Lammert. 2023. *Pseudo(-)* in French and Greek: Categorization and approximation. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1). 234–262.

Wheeler, Eric S. 2005. Multidimensional scaling for linguists. In Reinhard Köhler, Gabriel Altmann & Rajmund G. Piotrowski (eds.), *Quantitative Linguistics*, 548–553. (HSK 27). Berlin & New York: De Gruyter.

Bert Cappelle

Research Unit UMR 8163 STL, Univ. de Lille and CNRS, France

Domaine Universitaire du Pont de Bois

Rue du Barreau – BP 60149

59653 Villeneuve d'Ascq, France

bert.cappelle@univ-lille.fr

Robert Daugs

Christian Albrechts Universität zu Kiel

Englisches Seminar der CAU zu Kiel

Olshausenstraße 40

24098 Kiel, Germany

daugs@anglistik.uni-kiel.de

Stefan Hartmann

Heinrich Heine Universität Düsseldorf

Abteilung für Germanistische Sprachwissenschaft

Universitätsstraße 1

40225 Düsseldorf, Germany

hartmast@hhu.de

Matthias Eitelmann & Dagmar Haumann

Getting *close-ish*: A corpus-based exploration of *-ish* as a marker of approximation and vagueness

Abstract: This paper investigates the approximative nature of *-ish*, which takes its origin in de-adjectival adjectives in Middle English and from there spreads to *-ish*-derivatives from a wide array of bases, in terms of both categoriality and complexity. Drawing on data from the TV corpus, the paper charts the inventory of *-ish*-derivatives expressing approximative senses and zooms in on de-adjectival derivatives (*largish, small-er-ish*), de-numeral derivatives (*fourth-ish, 2.8-ish*), as well as on non-category changing *-ish*-formations (*a few weeks-ish*, *nothingish*). Building on both the diachronic trajectory of approximative *-ish* as of Middle English and the inventory of *-ish*-formations in the TV corpus, the paper proposes an approximation cline ranging from the earliest relational/associative senses of the suffix via similitudinal and genuinely approximative senses to incipient privative senses, (almost) all of which are attested in the data investigated.

Keywords: approximation cline, corpus-based analysis, derivation, frequency, *-ish* suffix, productivity, TV data

1. Introduction

Compared with its cognates in various other Germanic languages, the English derivational suffix *-ish* proves to be quite exceptional not only in that it may attach to a multitude of bases[1], but also in that it has developed approximative semantics in addition to the relational or associative senses that it has in common with, e.g., German *-isch*, Norwegian *-(i)sk* or Dutch *-s*. While the relational/associative senses manifest most prominently in nationality-denoting derivatives (1) as well as with denominal derivatives (2), the approximative sense, which is absent in all of the Germanic sister languages, can be seen in de-adjectival formations (3), numerals (4) or proper nouns (5).[2]

---

[1] The fact that *-ish* attaches to a wide categorial range of bases has been acknowledged among others by Marchand (1969: 305–306), Bauer et al. (2013: 311, passim), Dixon (2014: 96, passim) and Plag (2003: 96).

[2] Apart from the approximative function lacking in other Germanic languages, English *-ish* is also exclusive in that it has been extended to phrasal bases and may occur as a free lexical item.

(1)   Span*ish*, cf. German 'span*isch*', Norwegian 'span*sk*', Dutch 'spaan*s*'

(2)   heaven*ish*, cf. German 'himml*isch*', Norwegian 'himmel*sk*', Dutch 'hemel*s*'

(3)   clean*ish*, green*ish*

(4)   1-*ish*, 50-*ish*

(5)   Al Capon*ish*, James Dean*ish*

The crucial difference between the relational/associative meaning and the approximative meaning is that the former indicates a relationship of belongingness and similarity in the sense of 'characteristic of', whereas the latter ultimately emphasises a dissimilarity to a greater or lesser extent, ranging from 'vaguely x' to 'almost x, but ultimately not belonging to x' (on this issue see also Kuzmack 2007). This tension is humorously exploited in the following example taken from the TV corpus:

(6)   I discover I'm an eighth *Swedish*, which makes me *Swede-ish*. (2017, *The Real O'Neils*)

The utterance juxtaposes two diametrically opposed senses of *Swedish*: on the one hand, *Swedish* in the well-established meaning of an ethnic adjective, and on the other hand, a re-segmented derivative, with a hiatus between the base *Swede* and a phonologically stressed affix, that puts emphasis on the non-prototypical reading of *Swede-ish* as 'not Swedish, only vaguely so'. In this regard, the approximative -*ish*-derivative also denotes some sort of deficiency of a property.

The TV corpus abounds with instances similar to the one in (6), which indicate a meta-awareness of the approximative semantics of -*ish*-derivatives. Consider (7)–(9):

(7)   I wouldn't call it rude. Rude-*ish*. Rude-esque. Whatever you wanna call it. (2000, *Buffy the Vampire Slayer*)

(8)   Are you okay with the guy thing? – Yes. – Really? – Okay-*ish*. – That's not okay. – That's okay with an "ish". (2000, *Gilmore Girls*)

(9)   We're not platonic. Platon-*ish*, maybe. (2013, *How I Met Your Mother*)

In (7), the speaker explicitly contemplates the approximative nature of *rudish* vis-à-vis the positive form of the adjective, while at the same time offering another morphological means of marking approximation with the word-form *rude-esque*. (8) discusses the vagueness of *okay-ish* as opposed to the more affirmative *okay*, with an additional metalinguistic comment on the underlying morphological process. In (9), the suffix change from -*ic* to -*ish* results in

platon-ish as the opposite of *platonic*, which playfully conveys the notion that being in a 'platon-ish' relationship will most surely undermine the ethics of a platonic connection.

Examples like these raise a number of issues, which form the basis of two major research interests pursued in this paper. The most straightforward one concerns the task of generally taking stock of approximative morphology via *-ish*-derivation, thus going beyond those *-ish*-derivatives that meet the eye due to their comical quality. The inventory of *-ish* derivatives provides the opportunity for a closer inspection of those base categories where approximative senses are given most prominently. A second issue involves the differentiation of pragmatic uses and functions of approximative *-ish*-constructs. An exploration into the morphology-pragmatics interface on the basis of attested data allows us to perceive approximation as a more fine-grained notion and to establish a taxonomy of various approximative senses.

The paper is structured as follows. Section 2 revisits the evolution of approximative *-ish* in the history of English and elaborates on the impact that de-adjectival *-ish*-derivatives, which were the first ones to adopt the sense of approximation, had on the overall system of *-ish*-derivation. Section 3 then proceeds to discuss data from the TV corpus, starting out with a bird's eye view of the patterns of *-ish*-derivation that emerge from the corpus data (3.1.). The further analysis then zooms in on more fine-grained aspects pertaining to approximative senses, namely de-adjectival (3.2.) and de-numeral derivatives (3.3.) and a number of cases amongst hapax legomena constructs where *-ish*-suffixation does not yield adjectival outputs (3.4.), which seems to be a recent development tying in with other *-ish*-derivatives that do not induce a change of the base category. Section 4, then, recapitulates the approximative nature of *-ish*-derivatives by categorising them in terms of an approximation cline. The paper is wrapped up by a brief conclusion in section 5, which addresses to what extent the findings from the case study on the approximative nature of *-ish* bear a more general significance for the investigation of approximation in morphology.

2. Previously on (approximative) *-ish*

Approximative *-ish* as a morphological means to mark vagueness is an innovation of Middle English when the suffix came to be attached to adjectival bases – it is since then that the career of English *-ish* has parted ways with that of its Germanic cognates.

Before that pivotal moment, *-ish* had been a derivational affix exclusively conveying an associative sense that it shares with all sister languages, a meaning that is most prominently represented in derivatives denoting nationality/ethnicity and origin, such as Old English *englisc* 'English', *denisc* 'Danish' or *grecisc* 'Greek'. As the diachronic study conducted in Eitelmann, Haugland and Haumann (2020) shows, these ethnonymic expressions constitute the lion's share in terms of both types and tokens in the earliest period. Besides names for peoples as a reference to nationality or their respective language, this group of Old English *-ish*-derivatives also comprises names for cities (*lundenisc* 'from London'), countries (*ethiopisc* 'from Ethiopia') or individuals (*pontisc* 'of the Pontius family'), with the sense of relatedness being the common denominator in all of these cases. Apart from that, *-ish* also attaches to nominal bases – both simplex (e.g., *eotenisc* 'made by a giant' < *eoten* 'giant') and compound bases (e.g., *god-spellisc* 'evangelical' < *god-spell* 'gospel') –, once again conveying the sense of association firmly established with ethnic adjectives.

In Middle English, the pattern of *-ish*-derivation is affected by several changes, which set the scene for the ultimate extension of base categories beyond nouns and the evolution of the approximative sense. On the one hand, nominal bases have come to outnumber ethnonyms, which is due to a general decline of *-ish* as a nationality-denoting affix, ousted by other functionally equivalent affixes like *-ese* or *-ian* (cf. Dixon 2014: 268). At the same time, proper noun bases become virtually extinct. On the other hand, the associative sense with denominal derivatives increasingly shifts towards similitudinal meaning, expressing a likeness rather than an inherent relatedness (cf., e.g., *shepishse* 'sheep-like').

With de-adjectival derivatives entering the scene, *-ish*-derivation undergoes its most substantial change yet, namely the rise of approximative senses that Marchand (1969: 306) paraphrases as 'nearing, but not exactly x'. The widespread assumption that colour adjectives are the forerunners in this development (see, e.g., the OED entry on *-ish*, *suffix1*) could not be confirmed in Eitelmann, Haugland and Haumann (2020: 812), which drew on data from the *Middle English Dictionary* (MED): Formations based on colour adjectives such as *whitish* (1379), *yelwish* (1379) or *reddish* (1392) originate roughly at the same time as those based on common adjectives such as *fattish* (1369), *palish* (a1398) and *sourish* (a1398). Early metalinguistic comments found in medical treatises show that Middle English

writers were fully aware of a semantic difference between the common form of the adjective and its *-ish*-derived counterpart, as illustrated in (10):

(10)　　　Et for to know how Y tak þis term 'thenysshe,' tak hede þat þer is differens bytwene 'þenne' and 'thenn*ysshe*': for 'thenne' is propurly when it is mykel thenne (…); 'thenn*ysh*' when it is but o partie thenne, or elles menely thenne. (?c1400(1379) *Treat.Uroscopy [MED, s.v. *thinnish*])

'And in order to understand how I use this term 'thinnish,' note that there is a difference between 'thin' and 'thinnish': for 'thin' is proper when it is very thin (…); 'thinnish' when it is only partly thin, or else slightly thin.'

It is important to point out that in such early examples of de-adjectival *-ish*-derivatives, the meaning of attenuation is prevalent, with the explicit naming of the underived adjective serving as a baseline.

In Traugott and Trousdale's (2013) Construction Grammar analysis, the extension of *-ish* to adjectival bases is conceptualised as an instance of constructionalisation, i.e. the introduction of a new constructional schema with innovative formal and semantic properties that differ fundamentally from the ones already established in the constructional network. In other words, the admission of adjectival bases serves as the 'door opener' for approximative meanings which then also spread to other base categories. What we are dealing with in the case of *-ish* are two schemas (cf. Traugott & Trousdale 2013: 233–235). On the one hand, there is the original *-ish* schema attested since Old English times and shared by all Germanic languages, which has the meaning 'having character *of* x'. This schema is instantiated in two subschemas, i.e., an ethnic subschema and an associative subschema. On the other hand, the approximative *-ish* schema with the meaning 'having character *like* x', attested since Middle English, can be said to have developed from the original schema via the similitude sense ultimately resulting in a separate schema with its own distinct properties. Importantly, the rise of the approximative sense, initiated by the attachment of *-ish* to common adjectival bases, also has ripple effects for the meaning potential of denominal *-ish*-derivatives: while generally instantiating the sense of relatedness, denominal derivatives may occasionally also induce an approximative reading, particularly so as the meaning 'having the characteristics of x' is largely expressed by *-ish*'s competitor *-y*.

Once consolidated, the newly established approximative schema is affected by further constructional changes. Indeed, as of Early Modern English, there is a surge in *-ish*

constructs with an approximative meaning. At the same time *-ish*, generally, sees the laxing of selectional restrictions. As the array of bases that *-ish* attaches to becomes progressively more varied, both in terms of base category and complexity, approximative *-ish* constructs increase not only with adjectival bases, but also extend to numeral (*sixtyish*), adverbial (*foreverish*) or pronominal bases (*me-ish*). Numerals in particular serve to consolidate the approximative sense of *-ish*, together with proper noun bases that experience a comeback as of the 18th century. Resurfacing in Late Modern English, *-ish*-derivatives from names are no longer of the associative kind as in Old and Middle English, but now exclusively express approximation. For instance, *Graham Greenish* or *Londonish*, as attested in the British National Corpus (BNC), do not refer to something authored by Graham Greene or something directly related to London (as such formations would have in earlier periods); instead, they merely indicate a resemblance and simultaneously underscore an essential dissimilarity since *Graham Greenish* and *Londonish* imply that the entities thus described are ultimately *not* related to Graham Greene or London.[3]

A further crucial step in the development of *-ish*-derivation that contributed to the consolidation of the approximative sense concerns the extension to phrasal bases from the 18th century onwards so that *-ish* takes scope over an entire phrase, illustrated in (11) with examples from the Corpus of Contemporary American English (COCA):

(11) a. She finally issued just a sigh, but she was [[don't car]$_{TP}$ *ish*]$_A$ in her finery (…)

   b. What a saint! There was nothing remotely [dog-in-the-manger]$_{DP}$ *ish*]$_A$ about Beth, that was one of her special qualities.

Moreover, the attachment to complex phrasal units as in (12) may result in *-ish* displaying clitic behaviour, thus blurring the boundaries between morphology and syntax. Note that the clitic is less integrated than its affixal relative and can even constitute an intonation phrase of its own due to phonological strengthening (cf. Norde 2009: 224–225; Kuzmack 2007):

(12)   GEIST: How long have you been together?

---

[3] In this regard, German *-isch*-derivation is crucially different in that *-isch*-derivatives from names still indicate association, i.e., "das Shakespearsche Werk" 'the Shakespeare-ish work' is indeed authored by Shakespeare (cf. also Kempf & Eitelmann 2018 on this issue).

> KOTB: [[A year and a half]$_{DP}$ *-ish*]$_{DP}$. (COCA)

Although (11) and (12) are superficially similar, the approximative sense is only given in the latter case of clitic *-ish*. Also, the status of the respective *-ish*-formations is crucially different. Whereas clitic *-ish* in *a year and a half-ish* in (12) attaches to a phrase with no induced category change, *-ish* in *don't carish* and *dog-in-the-manger-ish* in (11) does derive an adjective and thus involves a phrasal or edge suffix.[4]

The continuous loosening of the selectional restrictions and the development of pure clitic *-ish* may have facilitated the rise of unhosted *ish* in the 20th century:

(13)   GIFFORD: So apparently we have two red wines in front of us [...]. All right, one of them is cheap...

KOTB: So – inexpensive.

GIFFORD: ... er, *ish*. (COCA)

The gradual debonding of *-ish*, to use Norde's (2009) term, is accompanied by a semantic shift from approximator to epistemic marker (cf. Traugott & Trousdale 2013: 236–237; see also Oltra-Massuet 2017: 312, passim).[5] Thus, closely related to the approximative senses as found with de-adjectival, de-numeral and deonymic bases, the free lexical item *ish* marks vagueness and, in addition, expresses speaker attitude, indicating a subjective assessment of the proposition. Traugott and Trousdale regard the rise of unhosted *ish* as another instance of constructionalisation since again, formal and semantic changes go hand in hand, with the extension to phrasal bases first facilitating cliticisation and then allowing for the occurrence of free *ish*.

Against the backdrop of this diachronically informed sketch of the evolution of approximative *-ish* vis-à-vis its associative counterpart, we will now embark on a further corpus-based inspection of *-ish*-derivatives in 20th/21st century English. In this way, we seek to gain a more fine-grained notion of what is commonly subsumed under the umbrella term 'approximative *-ish*' and thereby gain a deeper insight into approximation in morphology.

---

[4] We interpret Zwicky's (1987) original notion somewhat liberally (see Payne 2009 for a more recent implementation of edge morphology).
[5] But see Kuzmack (2007: 1, 8) and Norde (2009: 224–225) for a different view.

3. *-ish* in the TV corpus

3.1 Overview

In order to shed light on approximative *-ish* in contemporary English, we wanted to draw on a database that contains an informal variety of English as close to spoken speech as possible since we can assume approximative uses of *-ish* to thrive most prosperously in such contexts. To this end, we opted for the 325 million word *TV corpus* (www.english-corpora.org) as our data source, which comprises data from six English-speaking countries, covering a timespan from the 1950s to the 2010s.[6] This choice was made for primarily two reasons. On the one hand, various studies concerned with TV data have shown that scripted dialogue comes remarkably close to natural speech (see the discussion in Stange 2017). On the other hand, TV data can be expected to contain innovative language use that tests out the limits of established word-formation patterns, thus constituting a valid source to investigate language variation and ongoing changes.

Using *ish as our search string, we extracted all word- or phrase-final occurrences of *-ish*, with manually purging the strings to exclude any occurrences that end in the sequence *ish* but do not instantiate *-ish*-derivatives (e.g., *-ish* in verbs such as *establish* or in surnames such as *Whitish* or *Reddish*). This search string also allowed us to retrieve any instances of unhosted *ish* occurring in the data. In the case of spelling variants, relevant types of *-ish*-derivatives were subsumed under a unified lemma (e.g., *blonde-ish*, *blondish* and *blond-ish*). Subsequently, they were annotated for the category of the base *-ish* attaches to and the resultant category of the *-ish*-formation (if not adjectival).

We retrieved 73,717 *-ish* formations that fall into a total of 972 types, 618 of which are hapax legomena. Additionally, the data contain 402 cases of free lexical *ish*. The picture that emerges from the TV corpus snapshot is visualised in Fig. 1, which differentiates the respective distributions of the various base categories with respect to tokens, types and hapaxes. The miscellaneous category comprises pronouns (*nothing-ish*), prepositions (*through-ish*), interjections (*yeah-ish*) and degree markers (*kind of-ish*) serving as bases for *-ish*-attachment.

---

[6] See Davies (2020) on how the TV corpus was compiled drawing on freely available subtitles as well as for a well-argued rationale considering the use of TV data as an appropriate substitute for other spoken-language corpora, which are, by necessity, far smaller than the TV corpus.
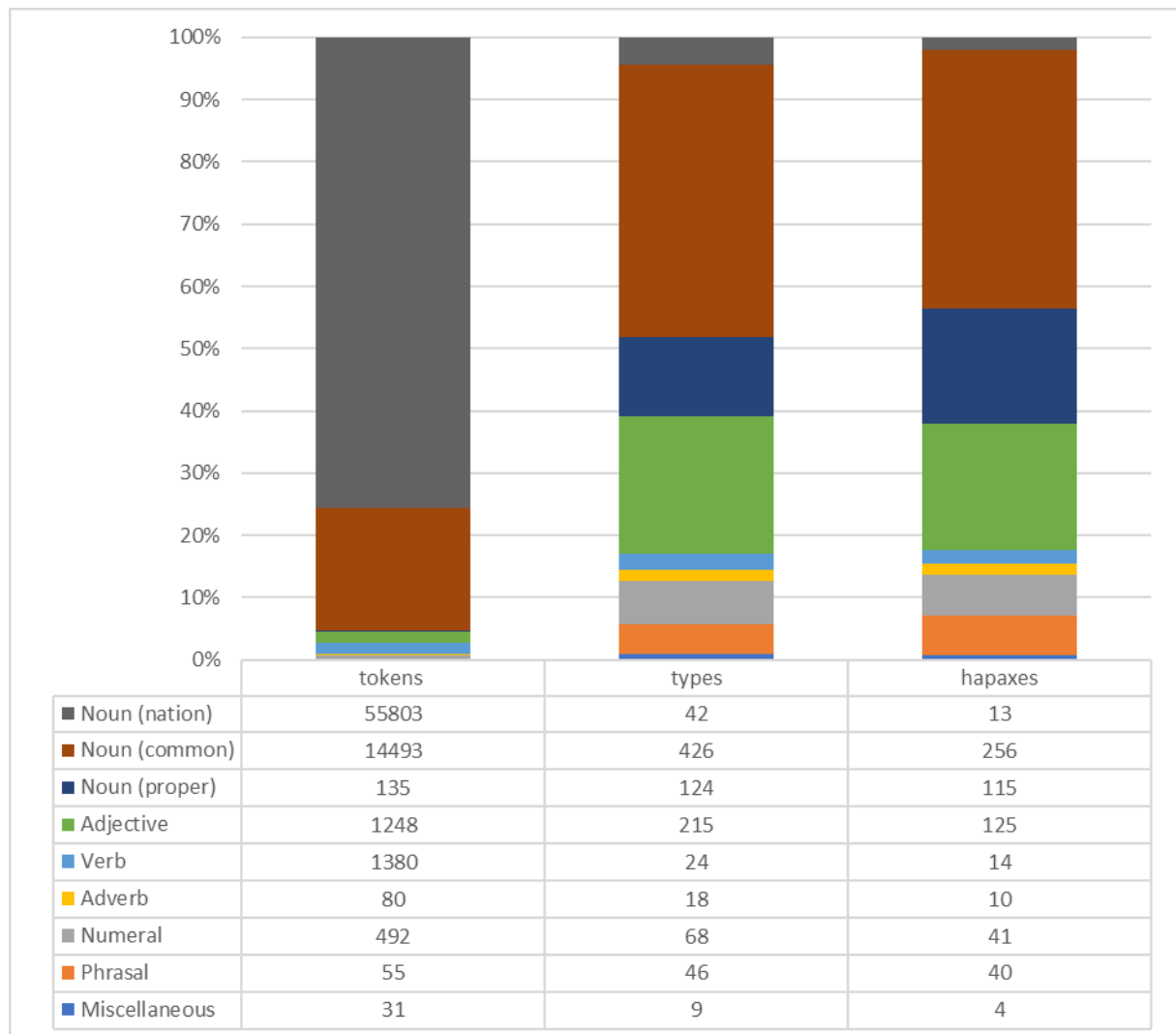
Fig. 1: Distribution of *-ish* formations in terms of tokens, types and hapaxes

	tokens	types	hapaxes
■ Noun (nation)	55803	42	13
■ Noun (common)	14493	426	256
■ Noun (proper)	135	124	115
■ Adjective	1248	215	125
■ Verb	1380	24	14
■ Adverb	80	18	10
■ Numeral	492	68	41
■ Phrasal	55	46	40
■ Miscellaneous	31	9	4

A differentiated glance at the individual base categories reveals where exactly *-ish* flourishes – and it is interesting to note that the order of tokens, types and hapaxes differs across all base categories. As Fig. 1 illustrates, nationality-denoting *-ish*-derivatives loom large token-wise, covering almost 75% of all tokens, but play an insignificant role with respect to types or hapaxes. This observation once again underscores that *-ish* no longer forms ethnic adjectives, with just a few exceptions such as *whaleish* (i.e. the language of whales) or *Dornish* (referring to the people from a place called Dorne in *Game of Thrones*), such attestations formed in analogy to the most frequent *-ish*-derivatives.

In terms of types and hapaxes, it is de-nominal and de-adjectival *-ish*-derivatives that feature prominently, taking first and second place, with 60% of de-nominal types and 58% of de-adjectival types being hapaxes. If we put the number of hapaxes in relation with the respective tokens, thus computing what Baayen (2009: 902) calls "potential productivity"

for individual base categories, we can assess "the probability of coming across new, unobserved types" (Plag, Dalton-Puffer & Baayen 1999: 215). In our case, every tenth de-adjectival *-ish*-derivative turns out to be a hapax, while this is true for only every hundredth denominal *-ish*-derivative. Since de-adjectival derivatives instantiate approximative meanings across the board, the approximative function of *-ish* can also gain a strong foothold in general. This assumption is supported by other base categories with a high hapax-type ratio, i.e., numerals (60%), adverbs (58%) or phrases (87%). Here, too, occurrences of *-ish* as a vagueness marker abound, which, again, points to the firm establishment of *-ish*'s approximative function:

(14)   I'd estimate the injury was sustained within the last three-*ish* weeks. (2017, *Silent Witness*)

(15)   How are you and Gunnar? Um... hanging on, sort of, barely*ish*. (2017, *Nashville*)

(16)   This is a man who believes in the rule of law-*ish*. The legislative process-*ish*. The Constitution-*ish*. (2015, *House of Cards*)

Note that in (16), the last two *-ish*-formations take scope over elided VPs, i.e., the VP *believe in* of the first sentence.

In the following, we are going to zoom in on those derivatives that express approximation and vagueness most conspicuously: de-adjectival (3.2) and de-numeral (3.3) formations. Also, we will briefly attend to some notable cases of approximative *-ish*-formations within the set of hapax legomena in which the attachment of approximative *-ish* does not yield an adjective as their outputs, thus not leading to a change of the base category (3.4).

3.2 Approximative *-ish* in de-adjectival formations

Claiming second place in terms of types and hapaxes, adjectival bases display a high degree of variability. Simplex qualitative adjectives figure prominently in *-ish*-formations (74%); these include relative adjectives (e.g., *stubbornish*, *cheapish*, *smallish*), gradable absolute adjectives (e.g., *cleanish*, *openish*, *fullish*) and non-scalar adjectives (e.g., *dead-ish*, *lesbianish*).[7] Morphologically complex formations represent more marginal types, with derived adjectives such as *beautifulish*, *educationalish* or *affordable-ish* (13%) outnumbering

---

[7] For finer grained distinctions within the class of absolute adjectives see, among others, Kennedy & McNally 2005; Sugawara 2012; Bochnak & Csipak 2014; Harris 2020. See also below.

compound adjectives such as *brand-new-ish* (3 out of 185).[8] Particularly noteworthy are de-verbal adjectives, e.g., *recommended-ish*, *intimidating-ish* or *drunk-ish* (9%) as well as four graded adjectives, i.e., 3 in the comparative (e.g., *laterish*) and one in the superlative (*bestish*). Formations from colour adjectives, in turn, first and foremost cover the usual suspects, i.e., *greenish*, *bluish* and the like (20 out of 30), with the colour hapaxes mostly covering compounded adjectives such as *orange-pinkish* or *grayish-yellowish* (9 out of 30). Note also *greeny-ish* for its sequence of two attenuating suffixes.

In order to check the syntactic function of de-adjectival *-ish*-derivatives, we took a closer look at the respective concordances. Leaving aside contexts in which de-adjectival *-ish* was not embedded in any kind or not unequivocally analysable, we found the majority of formations to be used attributively (65%),[9] followed by predicative uses (14%). Apart from that, a remarkable 12% of the contexts involve utterances in which speakers repeat a previously mentioned positive form of the adjective, thus adding a revised approximated perspective to the proposition, a context that is highly reminiscent of the earliest approximative uses discussed above in (10):

(17)   I'm not young. I'm not old-old. I'm just young*ish*. (1991, *Cheers*)

A closer look at the contexts also reveals to what extent de-adjectival *-ish*-derivatives may be premodified, which allows for a follow-up on an alleged restriction on *-ish*-formations as proposed by Morris (1998). While Morris concurs that *-ish*-derivatives from nominal bases "take intensifiers quite readily" (*very sheepish*, *truly fiendish*), *-ish*-derivatives from adjectives are said to "refuse most attempts at intensification" (Morris, 1998: 210). The non-availability of intensifiers with *-ish*-derivatives from adjectives (*?very oldish*, *?truly tallish*) is attributed to the derivative expressing attenuation/diminishment, which also purportedly precludes *-ish*-derivatives from comparatives/superlatives (**colderish*, **tallestish*) as these express "full lexical value[s]" (Morris, 1998: 210–211).[10]

---

[8] It should be noted that *intimidating-ish*, *affordable-ish*, *medievalish* etc. counter the oft-referred to tendency of *-ish* to combine with monosyllabic bases (cf. Dixon 2014: 235, Malkiel 1977: 348).

[9] We subsumed de-adjectival *-ish*-derivatives that premodify adjectives (i) under attributive:

(i) We saw a flashing *reddish* orange light coming through the pantry window. (Ancient Aliens, 2014)

[10] Apart from semantic reasons, there seems to be an additional morphologically related factor preventing formations such as **tallisher*, as Plag (2003: 177) elaborates: "[A]n easily decomposable suffix [such as *-ish*] inside a non-decomposable suffix [such as inflectional *-er*] would lead to difficulties in processing, whereas a less easily decomposable prefix inside a more easily decomposable suffix is easy to process." Therefore, the

As Morris' (1998) claims have been perpetuated in a number of publications, especially the claim that de-adjectival *-ish*-derivatives are essentially non-gradable adjectives and thus cannot co-occur with degree expressions or comparatives/superlatives (see, for example, Sugawara 2012: 10; Oltra-Massuet 2017: 63),[11] it is worthwhile testing whether these claims are corroborated empirically. We therefore coded the data for co-occurrences with all kinds of degree expressions ranging from garden-variety degree modifiers such as *a little (bit)* or *very* (18) over sundry adverbial modifiers, e.g., *relatively*, *fairly* (19) to comparatives and superlatives (20):

(18) a. You think you could sign this check for us? We're <u>a little bit </u>broke-*ish*. (2009, *Psych*)

b. Were they red? Like, <u>very</u> redd*ish*? (2016, *You Me Her*)

(19) a. For full disclosure's sake, I very much want it to end with <u>relatively</u> sober-*ish* sex. (2016, *You Me Her*)

b. I need a <u>fairly</u> larg*ish* malted milk. (1992, *Lovejoy*)

(20) a. Was he bigger than you? What? It's a valid question. Look, I'll give you he was small-*er*-*ish*. But quite mean. (2007, *Rules of Engagement*)

b. This new Mexican place just opened right by my parents' house. They have the <u>best-ish</u> tacos. (2008, *Brothers and Sisters*)

In addition, we coded for *kind of/kinda* and *sort of/sorta* (21) which, according to Oltra-Massuet (2017: 63), may co-occur with de-adjectival *-ish*-derivatives because they are approximators and not degree elements. However, given that *-ish* is an approximator itself, her argument loses force:

(21) a. I thought the shape was supposed to be <u>kind of</u> round*ish*. (2011, *Project Runway*)

b. it's my office, and I like to keep it <u>sort of </u>clean*ish*. (2005, *Criminal Minds*)

As these examples immediately illustrate, the claim that de-adjectival *-ish*-adjectives are non-gradable and resist degree modification is not supported empirically: de-adjectival *-ish*-formations do co-occur with a variety of degree expressions. We found a total of 152

---

sequence of affixes in *stronger-ish* or *small-er-ish* is accounted for, which still makes these forms appear unusual.

[11] According to Oltra-Massuet (2017: 57, passim), only *-ish*-derivatives from nouns and verbs are gradable; derivatives from adjectives, adverbs, numerals and phrases are not.

premodified de-adjectival *-ish*-derivatives. In other words, de-adjectival *-ish*-adjectives are gradable adjectives after all (pace Morris 1998; Sugawara 2012; Oltra-Massuet 2017).

The scale structural, i.e., gradable, properties of the adjectival base in *-ish*-derivatives have been shown to play a key role in the emergence of approximative readings. Thus, across frameworks and approaches, *-ish*-derivatives from relative adjectives which are associated with an open scale (e.g., *small*, *cheap*) as well as *-ish*-derivatives from gradable absolute adjectives which are associated with an upper-bound/upper-closed scale (e.g., *clean, safe*) are considered semantically well-formed because their scales do not have a default minimal standard. In other words, *-ish* targets a value that is slightly "below the standard for that adjective" (Bochnak & Csipak, 2014: 436; see, among others, Kennedy & McNally 2005; Sugawara 2012; Harris 2020 for discussion). Thus, *smallish* and *cheapish* in (22) and *cleanish* and *safe-ish* in (23) refer to a value that is (slightly) below the default value *small* and *clean*. In this respect, the derivatives in (22) and (23) express attenuation of the default (or standard of comparison) associated with the adjectival base:

(22) a. I'm looking for a young, single woman, driving alone in a small*ish* vehicle. (1990, *Inspector Morse*)

   b. And there's tons of storage. Yeah, and the rent's cheap*ish*, of course, compared to Montreal. (2009, *Being Erica*)

(23) a. There are clean*ish* towels in the bathroom. (2015, *The Middle*)

   b. He's one of the most modern doctors here at the historical dentists. You're in safe-*ish* hands. (2012, *Horrible Histories*)

Conversely, adjectives that are associated with a lower-bound/lower-closed scale (e.g., *open, dirty*) have a default minimum standard and thus do not sanction a degree lower than the lowest one on their semantic scale. However, as the examples in (24) illustrate, *-ish*-derivatives from these adjectives are felicitous.

(24) a. Do you have an open relationship? – Open*ish*. (2017, *When We Rise*)

   b. Yeah, it went from real blond to a dirty-*ish* blond in the 24 hours she's been with us. (2015, *Rosewood*)

Non-scalar adjectives (e.g., *dead*, *gay*) are not associated with scales and thus are non-gradable at all. Again, the *-ish*-derivatives *dead-ish* and *gay-ish* in (25) are felicitous:

(25) a. Although it does at least make her look dead-*ish*. (2013, *Face Off*)

b. This is the part where you say I'm the one who turned her gay-*ish*. (2017, *You Me Her*)

Accounts of semantically 'anomalous' but pragmatically felicitous *-ish*-derivatives from lower-bound/lower-closed and non-scalar adjectives build around the notion of 'imprecision', as captured by Morzycki's (2011) type shift operation PREC (originally postulated in connection with metalinguistic comparison), which makes available degrees of imprecision (see also Anderson 2016 for discussion). Another case in point is Lasersohn's (1999) Pragmatic Halos, which surround the truth-theoretic denotation of linguistic expressions and whose size is partly dependent on the presence of so-called 'slack regulators', such as *exactly* which requires a stricter interpretation, or *sorta* which allows for a looser interpretation (cf. Lasersohn 1999; Anderson 2016).

Sugawara (2012) and Bochnak and Csipak (2014), for example, argue that the imprecise uses of absolute adjectives with a lower-bound/lower-closed scale (cf. (23) above) result from a type-shifting operation by which the *-ish*-derivative targets "a scale of precision" rather than the bound/closed scale of its adjectival base (see Bochnak & Csipak 2014: 437, 440–441).[12] In their account, *-ish* is a 'precision regulator' giving rise to an approximative, 'around the threshold' interpretation. As has been argued by Anderson (2016: 45, passim) and Harris (2020: 79–81), for example, non-scalar adjectives can be coerced into gradable (absolute) adjectives and thus allow for pragmatic slack, as Anderson (2016: 17) illustrates with (26), to which we may add the examples in (25) above:

(26)   I've been *sorta* pregnant four times. Being *sorta* pregnant *sorta* sucks. It's like you're late, you test early, you see two lines, you go for a blood test, you're pregnant, and then it's JUST KIDDING!

In brief, it appears that deadjectival *-ish*-derivatives receive an attenuative or approximative interpretation. *-ish*-derivatives from relative adjectives and adjectives with upper-bound/upper-closed scales target an attenuated default, whereas *-ish*-derivatives from adjectives with lower-bound/lower-closed scales and non-scalar adjectives approximate a default (by targeting the pragmatic halo). In those cases in which deadjectival *-ish*-derivatives are modified, the modifier refers to degrees of attenuation or the size of the pragmatic halo.

---

[12] They do not discuss *-ish*-derivatives from non-scalar adjectives.

3.3 Approximative *-ish* in de-numeral derivatives

The natural numbers serving as bases for *-ish*-derivation attested in the TV corpus data are remarkably diverse, going beyond the kind of de-numeral bases discussed in Ruzaitė (2012); while she did not find certain numbers, e.g., 11, 80 or 90 to be suffixed by *-ish*, these interestingly do occur in our data. There are even some rare cases of suffixed ordinal numbers, which are particularly noteworthy in that the attachment of approximative *-ish* undermines their function of linearly ordered labelling:

(27)   You really think we shouldn't tell them we're fourth-*ish*? (2009, *Greek*)

(28)   ...you find a perpetrator you call me first before you put your hand to your head. – First? – First. – How about first-*ish*? (2010, *Psych*)

Yet, even if the amount of numbers that approximative *-ish* could possibly attach to is basically infinitesimal, the 68 de-numeral types in our dataset show clear preferences for high-frequency cardinal numbers from 1 to 10, amongst the teens the numbers 12 and 15, tens such as 20, 30, 40, 50 etc. and tens with the digit 5 such as 25, 35, 55. This observation might not come as too much of a surprise since these are the numbers typically used for calendar-clock references (29), to talk about a person's age (30) or to make a guess about an object's decade of origin (31) – communicative scenarios which in our data make up 87% of contexts in which de-numeral *-ish*-derivatives are used:

(29)   I'll come to you. Hotel bar, 7-*ish*? Can't wait. (2016, *Queen Sugar*)

(30)   I guess there must be a reason why women have babies when they're 20*ish* instead of 40*ish* (1989, *The Golden Girls*)

(31)   Bingo. Brass. Looks 1890s-*ish*, doesn't it? (2011, *Warehouse 13*)

In all these cases, the numbers thus approximated are not to be interpreted as exact values. Rather, they spell out Lasersohn's (1999) pragmatic halos and indicate pragmatic slack: the meeting time in (29) is not 7 sharp but is only near to 7 (plus/minus a few minutes), the women in question in (30) are not exactly 20 or 40 years of age but at an age when they can still be called to be in their 20s or 40s (which at the same time implies that they are probably not in their *late* 20s or 40s), and in (31), the object at hand is from any year belonging to the decade of the 1890s. It is crucial to note that such values allow for some interpretative tolerance since they involve numbers which refer to clear points of orientation, i.e., full hours, prototypical hours-minutes-combinations (mainly quarter to/past and half),

significant birthdays, memorable years, decades, etc. However, we also find quite a few numbers where such a threshold interpretation is not readily given, as for instance with some more random numbers as in (32) or (33) or with highly precise decimal numbers (34):

(32)   Boom. Married. That would make you... 32-*ish*. (1996, *Everybody Loves Raymond*)

(33)   Yep, we've had it since about 2005-*ish*. (2010, *Museum of Life*)

(34)   Cool. 3.4 pounds. That's heavy-*ish*. A female brain tips the scales at probably 2.8-*ish*. Average male brain weighs just over three pounds. (2011, *Crime Scene Investigation*)

While in (29)–(31) the suffixed numerals could be rendered as 'not exactly 7/20/40/1890 but sufficiently close to it', those in (32)–(34) do not seem to allow for a similar paraphrase. For instance, in (32), if marriage does not make someone exactly 32, the question is what *32-ish* is supposed to refer to;[13] those numbers coming close to 32 would be 31 or 33, but just like 32, they do not normally indicate any significant ages. Instead of giving an approximative reading to the base, such cases rather attenuate the speaker's commitment to the proposition, thus serving as downplaying hedging devices, which is also indicated by the conspicuous co-occurrence with further mitigating lexical items such as *about* or *probably* (on the issue of multiple hedging with approximated numbers see also Ruzaitė 2012: 195).

3.4 Non-category-changing -*ish*-derivatives

Closely related to de-numeral -*ish*-formations, which, in the majority of cases, do not derive adjectives but retain their numeral status, the TV corpus data contain quite a few further instances spotted amongst the hapax legomena, where the category of the base -*ish* attaches to remains unchanged, thus not yielding an adjectival output. This seems to be an innovation of the late 20th century; as far as we can tell from the diachronic study conducted in Eitelmann, Haugland and Haumann (2020), cases comparable to the ones discussed here are not attested in the periods before present-day English.[14] In this respect, they tie in with a development triggered by de-adjectival derivatives (which are also non-category changing even if they still result in an adjectival derivative), followed up by other derivatives

---

[13] An anonymous reviewer suggested that approximated precise numbers, e.g. (32), may have a surprisal (comic) value attached to them.

[14] The earliest example of a non-category changing -*ish*-derivative apart from numerals comes from the BNC in which de-adverbial *forever-ish* is still used adverbially:

(i) the forever-*ish* trickly sound of her high giggle

which further instantiate the approximative sense originally consolidated by de-adjectival derivatives, but which do not generate adjectival outputs.

The first example is an instance of a denominal derivative retaining its noun status:

(35)   So, when I met you on the bus, with your son
         – Son-*ish*. (2013, *Seed*)

Here, the protagonist Harry corrects the other speaker's false assumption that the child he was with was his actual son, the derivative *son-ish* alluding to the fact that he is only the kid's biological father (which is, all in all, a complicated constellation that the sit-com *Seed* humorously exploits). Verbs, too, can take *-ish* affixation without changing their category as verbs:

(36)   My ship automatically, uh, noticed-*ish* that your ship was having some bother. (2011, *Doctor Who*)

(37)   You have to know exactly? Couldn't you just know-*ish*? (1998, *The Practice*)

The same holds true for adverbs (38), prepositions (39) and pronouns (40) the latter of which are particularly noteworthy since pronouns are functional items and as such are not expected to participate in word-formation processes.

(38)   Where did you leave her? Um, uh - - you know, there-*ish*. (2012, *The Big C*)

(39)   You sure you can get us around that shield? Well, not around, but through-*ish*. As in, between the molecules. (2017, *Guardians of the Galaxy*)

(40)   Nothing happened. Nothing*ish*. (2015, *The Mysteries of Laura*)

Finally, we also find a number of affixed phrases which preserve their phrasal status; these involve cases of clitic *-ish*. In (41), *-ish* combines with the entire NP and thus approximates the referent of the NP, a measured-out time interval. Note that *-ish* follows the inflectional suffix, which is indicative of *-ish* attaching to the entire phrase. In (42), *-ish* combines with the entire VP, yielding an approximated reading of 'seeing someone', and in (43), with the PP *four years ago*, *-ish* approximates the timespan that has elapsed between the time of the event and the time of speech.

(41)   What, for a few days, or...? I was thinking more like a few weeks-*ish*. (2011, *Heartland*)

(42)   I'm kind of seeing someone-*ish*. (2016, *2 Broke Girls*)

(43)   When? – Uh, four years ago-*ish*. (2014, *Rizzoli & Isles*)

The fact that these 'new' constructs differ from 'regular' *-ish* derivatives in that the category of the base remains unchanged and that their meaning is unequivocally approximative makes them reminiscent of Scalise's (1986: 131ff., 1988: 233ff.) notion of evaluative affixes. De-numeral derivatives may be assumed to have triggered the increasing consolidation of non-category-changing derivatives: Just like de-adjectival derivatives paved the way for the approximative meaning with further base categories, the widespread use of approximative numbers with no induced category change might have opened the door for similar uses in other contexts as well. In any case, what we can observe here is a new constructional change in the making.

4. *-ish*, approximately

Our analysis of de-adjectival and de-numeral *-ish*-derivatives as well as the exploration of non-category changing formations as attested in the TV corpus has shown that derivatives instantiating what is commonly referred to as 'approximative *-ish*' actually display subtly different nuances of meaning. Ultimately, the label 'approximative *-ish*' rather serves as an umbrella term that covers quite a wide range of different senses, with the notion of approximation being more or less evident. We can thus use the data to establish a taxonomy of approximation that also takes into account the extent to which approximation is related to the expression of inherent relatedness, similitude and attenuation, particularly so as these senses are intricately interwoven as shown by the trajectory of change which the pattern of *-ish* derivation has undergone.

Against this backdrop, we can conceive of approximation as a gradient ranging from relation/association at the one end of the scale to approximation/privation at the other end of the spectrum, the two endpoints representing the contrast between inherent similarity and ultimate dissimilarity. We thus suggest the approximation cline in Fig. 2, which may be read both diachronically and synchronically. To spell this out, on the one hand, it regards the different grades of approximation as evolving from each other in a certain

chronological order.[15] On the other hand, it displays the wide array of different meanings that the pattern of *-ish*-derivation instantiates from a synchronic point of view. It needs to be kept in mind that *-ish* is quite a special case in that it manifests almost all stages in the cline since older and newer functions co-exist.

relational > associative > similitudinal > attenuative > approximative > disproximative > privative

Fig. 2: Approximation cline

As pointed out in the diachronic overview (see section 2), fully relational senses are instantiated by nationality-denoting *-ish*-derivatives: these well-established ethnic adjectives, which make up the largest share of tokens in our data, indicate an intrinsic relationship to the name of nations or ethnic groups. With a few exceptions (see section 3.1), novel deonymic formations are of the approximative kind as, e.g., *Latino-ish* or *Norwegish*.

Similarly to such relational senses with ethnic adjectives, derivatives realizing associative senses constitute an identity relation to the base, which can be paraphrased by 'characteristic of/typical of'. However, purely associative *-ish* has largely given way to the expression of similitudinal senses, such derivatives emphasising the likeness to the entity denoted by the noun they attach to. Put differently, while derivatives of the associative kind are intrinsically connected to the base, derivatives of the similitudinal kind imply a comparison, thereby focussing on either particularly salient features of the base or the entire category (cf. Bauer, Lieber & Plag 2013: 312). These two readings are nicely illustrated by the following two examples:

(44)   Monkey, you're all limber and bendy and monkey-*ish*. (2012, *Kung Fu Panda*)

(45)   … Finlay is the small, well-intentioned monkey-*ish* little fellow. (2013, *Misfits*)

Used in connection with *limber* and *bendy*, *monkey-ish* in (44) characterises the monkey's behaviour as typical of a monkey, whereas *monkey-ish* in (45) merely states that Finlay's

---

[15] The assumption that approximation results from similarity by way of inference is also discussed by Bauer, Lieber and Plag (2013: 313) for the case of de-adjectival derivatives: "If we say something is similar to [qualities like] *dull*, (…), the inference is drawn that we cannot mean exactly *dull* (…) but rather must mean something not exactly the same as those qualities, that is, approximating those qualities."

appearance reminds one of a monkey (without insinuating, of course, that he *is* a monkey). A cursory look at new de-nominal formations reveals that most of them express a similarity sense, with the associative sense rather being conveyed by competitor *-y*. At the same time, we find similitudinal *-ish* competing with suffixes like *-like* or *-esque*, as illustrated in (46):

(46)   Hey. Is everything in here onion-like, bacon-esque or cheese-*ish*? (2010, *Desperate Housewives*)

Similarity is just one step away from the expression of attenuation, i.e., the reduced degree of a quality, which then is prototypically instantiated by de-adjectival formations from colour adjectives: *greenish*, for example, is just a weaker form of *green*, albeit still essentially green. Such a sense of reduction is also given with other de-adjectival derivatives: thus, *-ish*-derivatives from relative adjectives and gradable adjectives with an upper-bound/upper-closed scale, as in (47) and (48), can be paraphrased as 'weakly ADJ' or 'vaguely ADJ':

(47)   A simple girl from a one-stop-light town, and you, the worldly, handsome-*ish* Sam Winchester. (2016, *Supernatural*)

(48)   The great and the good clamouring for invitations to tea. Well, the *greatish* and *good-ish*, anyway. (2015, *Life in Squares*)

Unequivocally approximative senses are evidenced in those cases where the resultant derivative's meaning revolves around a default threshold value, i.e., 'nearing X' or 'almost X'. In other words, we have moved further down the approximation cline towards the dissimilarity end. As elaborated above, de-adjectival formations from both gradable adjectives with a lower-bound/lower-closed scale (49) and non-scalar adjectives (50) as well as numerals (51) play an important role in this context. Note also (52), which adds approximative *-ish* to relational *-ish*:

(49)   Do you have an open relationship? – Open*ish*. (2017, *When We Rise*)

(50)   Um, it just seems a little dead. – He's not dead. Dead-*ish*. (2016, *Rizzoli & Isles*)

(51)   At that hour? – Yeah, it's open till, like, ten-*ish*. (2017, *Gypsy*)

(52)   You've been spending a lot of time with Rita, and she's British-*ish*. (2005, *Arrested Development*)

Derivatives from names are particularly noteworthy in this context since their semantics is in-between similitudinal and approximative, with a strong sense of subjectivity.

(53) a. Betsy, this is hot. This is actually like lady-gaga-*ish*. (2013, *The Mindy Project*)

b. He's gonna act all normal and un-Norman-Bates-*ish* and then once I drop my guard, bam, that's when he strikes. (2012, *One Tree Hill*)

In all these cases, if the *-ish*-derived name derivatives were replaced by simply the name without a suffix or by adding *like*, these would indicate the comparison explicitly, thus implying a direct comparison to the person in question. In contrast, the *-ish*-derived variants abstract the person referred to by metonymically approximating their most salient characteristics (Lady Gaga's shrill outfits in (53a)) or their persona as a whole (Norman Bates' psychopathic behaviour in (53b)).

Disproximation in the sense of Cappelle, Daugs and Hartmann (2023) obtains in those cases where the resultant derivative denotes that a standard is not met, thus refuting a prototype reading – as opposed to approximation proper, which expresses that something comes close to a standard or baseline level. The non-category-changing formations in, e.g., (35) and (39) above fit into this class in that a [son-*ish*]$_N$ is not a son proper, or [through-*ish*]$_P$ does not describe a straightforward movement through something. In this regard, such derivatives have a downscaling effect, which is also observable with the following non-category changing de-adverbial formations:

(54) a. So we've got a highly-*ish* skilled covert agent who probably has millions of dollars in cash. (2011, *Archer*)

b. Maybe. – Maybe? – Not a definite maybe. A maybe-*ish* maybe. (2013, *Raising Hope*)

While in (54a) the intensifying effect of *highly* is turned down a notch, thus making "a highly-*ish* skilled covert agent" essentially 'a not entirely highly skilled covert agent', the inherent vagueness of *maybe* is made even more vague in (54b), with *maybe-ish* an even weaker commitment than *maybe*. For good measure, we can add here the examples in (55), where *-ish* enhances the hedging or epistemic stance-marking function of *kind of/sort of*:

(55) a. Weren't you, like, antisocial, kind of-*ish*? (2015, *I am Jazz*)

b. Absolutely! Almost completely positive, sort of-*ish*. (2016, *Teenage Mutant Ninja Turtles*)

The sense of disproximation also comes into play in cases of free *ish* as hedges or epistemic stance markers which downplay the speaker's commitment to the truth value of the proposition:

(56) a. I can cook. *Ish*. (2011, *Supernatural*)

b. There's a black market for flowers? *Ish*. Most transactions are completed via online auction sites. (2015, *Elementary*)

The ultimate end of the approximation scale is reached with fully privative senses 'not X', which privative prefixes such as *fake-* and *pseudo-* express (cf. Van Goethem & Norde 2020). *-ish*, as of yet, does not seem to realise such privative meanings – and still, rare findings such as *friendship-ish* in (57) seem to be a close match since the relationship that the speaker suggests to forge would only pretend that it is friendship:

(57) So we're, you know, forging this friendship-*ish* type thing? (2012, *Jane by Design*)

The examples presented in this section illustrate the intricate nature of approximative *-ish*-derivatives: rather than constituting one single sense, approximation is a gradient between inherent similarity, as expressed by relational *-ish*-derivatives, and implied dissimilarity, as expressed by disproximative *-ish*-derivatives.

5. Conclusion

This paper investigated the English suffix *-ish* as a morphological means to convey approximation and vagueness with two main objectives. The first concerned providing an inventory of *-ish*-derivatives in order to inspect more closely those base categories related to approximative senses. Against the backdrop of *-ish*'s trajectory of change particularly since Middle English and on the basis of data from the TV corpus, we first of all took stock of the base categories that *-ish* attaches to: while ethnic adjectives make up the lion's share in the data token-wise, similitudinal and approximative *-ish*-formations abound type-wise. We then revisited those base categories that lend themselves most readily to approximation, i.e., adjectival and numeral bases, and spotted some ongoing changes concerning *-ish* formations that do not induce a change of base category and thus do not yield adjectival outputs. As *-ish* continues to be used innovatively and creatively, it will definitely be interesting to keep track of any further developments. It remains to be seen whether *-ish* can hold its ground as a multifunctional affix with a further expansion of its cross-categorial nature, or whether one of the other adjective-deriving suffixes such as *-y* as a close competitor will

catch up in the long run, potentially restricting *-ish*'s range of application to predominantly the approximative function in novel word-formations.

The second objective was to establish a taxonomy of approximative senses. Following up on the assumption that approximation results from association, we proposed an approximation cline that goes from relational senses on the one end of the scale to privative senses on the other end, with an intrinsic identity relation successively transgressing into an ultimate non-identity relation via the intermediary steps of similitude and approximation proper. As we saw, *-ish*-formations as attested in the TV corpus realise (almost) all grades of the approximation cline, which further highlights the versatility and multifunctionality of this suffix.

Going beyond the case study on *-ish*, the proposed approximation cline bears general implications for the investigation of approximation in morphology. We can expect morphological means of expressing approximation to develop out of derivational morphology originally denoting associative senses. Whether adjectives act as a door opener for the establishment of approximative senses is something that remains to be investigated more thoroughly; it is conspicuous, however, that the English derivational affix *-y* also displays approximative semantics (albeit weakly in comparison to *-ish*) when attached to adjectives as, e.g., in *cheapy* (vs. *cheap*) or *greeny* (vs. *green*). Also, it is of interest to test the cross-linguistic validity of the proposed approximation cline, i.e., to investigate to what extent a given derivational affix that has developed approximative functions along the cline loses its former, less approximative senses. As the case study on *-ish* shows, synchronically, all of its functions, both original and more recent, are readily attested in a scenario of layering (Hopper 1991: 22), which may not necessarily be the case for other affixes that have moved down the cline.

References

Anderson, Curt. 2013. Inherent and coerced gradability across categories: Manipulating pragmatic halos with *sorta*. In Todd Snider (ed.), *Proceedings of the 23rd Semantics and Linguistic Theory Conference* 23. 81–96.

Anderson, Curt. 2016. *Intensification and Attenuation across Categories*. PhD dissertation. Michigan State University.

Baayen, Harald. 2009. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling & Merja Kytö (eds.), *Corpus Linguistics. An International Handbook.* Vol. 2, 899–919. Berlin: Mouton de Gruyter.

Bochnak, M. Ryan. & Csipak, Eva. 2014. A new metalinguistic degree morpheme. In Todd Snider, Sarah D'Antonio & Mia Weigand (eds.), *Proceedings of the 24th Semantics and Linguistic Theory Conference*, 432–452.

Davies, Mark. 2019. *The TV corpus.* https://www.english-corpora.org (accessed 9 March 2023).

Davies, Mark. 2020. The TV and Movies corpora. Design, construction, and use. *International Journal of Corpus Linguistics* 26(1). 10–37.

Dixon, R. M. W. 2014. *Making New Words. Morphological Derivation in English*. Oxford: Oxford University Press.

Eitelmann, Matthias, Kari E. Haugland & Dagmar Haumann. 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.

Harris, Tabea. 2020. Vagueness, context-sensitivity and scale structure of four types of adjectives with the suffix *-ish*. In Remus Gergel & Jonathan Watkins (eds.), *Quantification and Scales in Change*, 67–84. Berlin: Language Science Press.

Hopper, Paul J. 1991. On some principles of grammaticization. In Elizabeth C. Traugott & Bernd Heine (eds.), *Approaches to Grammaticalization. Volume 1: Focus on Theoretical and Methodological Issues*, 17–35. Amsterdam: Benjamins.

Kempf, Luise & Matthias Eitelmann. 2018. Von *diutisk* zu *dynamisch*, von *englisc* zu *anything-ish*. *-isch* kontrastiv diachron. *Zeitschrift für Wortbildung/Journal of Word Formation* 2(1). 93–134.

Kennedy, Christopher & Louise McNally 2005. Scale structure, degree modification and the semantics of gradable predicates. *Language* 81(2). 345–338.

Kuzmack, Stefanie. 2007. Ish: A new case of antigrammaticalization. Paper presented at the Annual Meeting of the Linguistic Society of America, Anaheim, CA, 4–7 January.

Lasersohn, Peter. 1999. Pragmatic halos. *Language* 75(3). 522–551.

Malkiel, Yakov. 1977. Why Ap-ish but Worm-y? In Paul Hopper (ed.), *Studies in Descriptive and Historical Linguistics. Festschrift für Winfried P. Lehmann*, 341–364. Amsterdam & Philadelphia: Benjamins.

Marchand, Hans. 1969. *The Categories and Types of English Word-formation*. 2nd ed. Munich: Beck.

Morris, Lori. 1998. A toughish problem: the meaning of *-ish*. *LACUS Forum* 24. 207–215.

Morzycki, Marcin. 2011. Metalinguistic comparison in alternative semantics for imprecision. *Natural Language Semantics* 19. 39–86.

Norde, Muriel. 2009. *Degrammaticalization*. Oxford: Oxford University Press.

Oltra-Massuet, Isabel. 2016. Propositional *ish* as a syntactic Speech Act Phrase. In Aitor Ibarrola-Armendariz & Jon Ortiz de Urbina Arruabarrena (eds.), *On the Move: Glancing Backwards to Build a Future in English Studies*, 307–313. Bilbao: University of Deusto.

Oltra-Massuet, Isabel. 2017. Towards a morphosyntactic analysis of *-ish*. *Word Structure* 10(1). 54–78.

Payne, John. 2009. The English genitive and double case. *Transactions of the Philological Society* 107. 322–357.

Plag, Ingo. 2003. *Word-formation in English*. Cambridge: Cambridge University Press.

Plag, Ingo, Christiane Dalton-Puffer & Harald Baayen. 1999. Morphological productivity across speech and writing. *English Language and Linguistics* 3(2). 209–228.

Ruzaitė, Jūratė. 2012. Attenuative morphological forms across varieties and registers: a corpus-based study of the suffix -ish. *Deeds and Days* (Darbai ir Dienos) 58. 183–200.

Scalise, Sergio. 1986. *Generative Morphology*. Dordrecht: Foris.

Scalise, Sergio. 1988. The notion of 'head' in morphology. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology*, 229–245. Dordrecht: Foris.

Stange, Ulrike. 2017. 'You're so not going to believe this': The use of GenX *so* in constructions with future *going to* in American English. *American Speech* 92(4). 487–524.

Sugawara, Ayaka. 2012. *Semantics of English suffix* -ish. Paper Presented at the 48th Annual Meeting of the Chicago Linguistic Society, Chicago, MI, 20 April 2012.

Traugott, Elizabeth C. & Graeme Trousdale. 2013. *Constructionalization and Constructional Changes*. Oxford: Oxford University Press.

Van Goethem, Kristel & Muriel Norde. 2020. Extravagant "fake" morphemes in Dutch. Morphological productivity, semantic profiles and categorical flexibility. *Corpus Linguistics and Linguistic Theory* 16(3). 425–458.

Zwicky, Arnold M. 1987. Suppressing the Z's. *Journal of Linguistics* 23(1). 133–148.

Matthias Eitelmann
Johannes Gutenberg-University Mainz
English Linguistics
Jakob-Welder-Weg 18
55128 Mainz, Germany
eitelman@uni-mainz.de

Dagmar Haumann
University of Bergen
Department of Foreign Languages
Postbox 7805
5020 Bergen, Norway
dagmar.haumann@uib.no

Matthias Hüning & Barbara Schlücker

Approximation and comparison in word-formation: The case of denominal adjectives in Dutch, German, and English

Abstract: The paper presents a corpus-based comparative study of denominal adjectives in Dutch, German, and English. It aims at clarifying the notion of approximation. More specifically, it focuses on investigating the relation between approximation and comparison as semantic categories in word-formation. Drawing on equivalent patterns, we study both the relation between comparative and approximative readings in various word-formation patterns (derivation and compounding) in Dutch, German, and English as well as the respective differences between the languages. We specifically focus on suffixes and other right head constituents (in particular Dutch *-achtig*, German *-artig* and *-ähnlich*, and English *-like*), thereby addressing the question whether word class changing morphology is at all suitable for expressing approximation. We conclude that the patterns investigated in our study may have approximative readings but are not approximative in the strict sense.

Keywords: comparative morphology, approximation, semantics of word-formation, compounding, derivation, evaluative morphology, contrastive linguistics, West Germanic

1. Introduction[*]

Approximation is often paraphrased as approaching or imitating a concept. Typical examples of approximative morphology from the literature (in the domain of nominal and adjectival word-formation) are compounds with *pseudo-* or *semi-* (*pseudo problem*, *pseudo-transitive*, *semi-professional*). In traditional word-formation theory (on German, see e.g., Kühnhold et al. 1978: 188ff), these examples are semantically classified as privative word-formation. In such a semantic categorization, privative morphology stands next to other semantic groups such as comparison or gradation.

Studies on approximation often indicate a connection or overlap with evaluative morphology. For instance, some authors consider approximation as a sub-domain of evaluation

---

[*] We would like to thank the anonymous reviewers for their valuable and constructive comments, which certainly helped to sharpen our argument and improve the paper.

and approximative morphological markers as evaluative markers from a functional point of view (e.g., Masini & Micheli 2020; Van Goethem & Norde 2020). Evaluative patterns serve to express speaker attitude and evaluation. Morphologically, they can be realized in different ways, mainly by affixes (e.g., German ***Ge****singe* (pejorative term for) 'singing', *Min-**isterchen*** lit. little minister) and affixoids (***sau****kalt* lit. sow cold, 'freezing cold', ***grotten-**schlecht* lit. cave bad 'lousy').

From a semantic point of view, approximation (as a case of privative morphology) also resembles comparison. This resemblance is especially strong in adjectival word-formation (e.g., Kühnhold et al. 1978; Fleischer & Barz 2012) (German *heldenhaft* 'heroic', *hündisch* 'doglike', *grippös* 'influenzal', *grippoid* 'grippoid', *engelsgleich* 'angelic', *menschenähnlich* 'human-like'). While evaluation is quite clearly a functional category in its own right, the semantic demarcation between approximation and comparison is much more difficult. In particular, there is no clear and exclusive assignment of morphological markers to either of these categories. This can be illustrated by the Dutch suffix *-achtig*. A formation like *leerachtig* (*een leerachtige substantie* 'a leathery/leather-like substance') might be classified as approximative if it is intended to express that the substance has leather-like properties but is not leather, i.e., when it is approximating the category leather (cf. (1)). On the other hand, it can be classified as comparative, if it is intended to refer to the presence of leather-typical properties, regardless of whether it is actually leather or not (cf. (2)). It is not always easy to differentiate between the two; both meanings seem to be possible in the case of (3):

(1) Het kunstleder is een goed alternatief voor de stoffen en nylons. Het materiaal voelt zacht en **leerachtig** aan. (nlTenTen14)
'Faux leather is a good alternative to fabrics and nylons. The material feels soft and leather-like.'

(2) Skimmea Japonica is een groenblijvend heestertje dat niet hoger dan een meter wordt met een zeer compacte groei en **leerachtig** ovaal blad. (nlTenTen14)
'Skimmea Japonica is an evergreen shrub that does not grow higher than one meter with a very compact growth and leathery oval leaves.'

(3) Afgelopen zaterdag verdwenen uit de dameskleedkamer: bruin **leerachtig** kinder-jasje. (nlTenTen14)
'Disappeared from the ladies' locker room last Saturday: brown leathery/leather-like children's jacket.'

Approximation and comparison are both related to the concept of similarity. However, although similarity (or resemblance) is an integral part of semantic descriptions of approximation in the literature, the distinction and demarcation of approximation and comparison has hardly been discussed so far.

For this reason, our contribution aims at describing and defining approximation more precisely as a category in its own right by comparing and distinguishing it from comparison as a central semantic group in word-formation. We argue that approximation should neither be equated with comparison nor similarity and that similarity is a necessary but not a sufficient part of the concept of approximation in the sense that approximation is derived from similarity through inference (Bauer et al. 2013: 313).[1] In other words, the expression of similarity does not suffice for qualifying as approximation. At the same time, however, our study also shows that the conceptual distinction between approximation and comparison cannot always be clearly made for all markers under discussion. We will, therefore, argue that while there are morphological markers that are unambiguously and exclusively approximative there are also other markers that are basically comparative but may have approximative readings in certain contexts.

In the following, we will focus primarily on such markers. We follow a contrastive approach, dealing with denominal adjective formation (derivation and compounding) in German, Dutch and English. The markers examined here are particularly interesting since all of them are right constituents (suffixes and stems) and thus heads that determine the word class of the resulting formations. In contrast, many of the known examples of approximative morphology from the literature are left constituents, namely prefixes, prefixoids, confixes, and stems. These elements are not word class changing since the word class is determined by the right constituent of the complex word.

Finally, our investigation also touches on issues of competition of morphological patterns, since it involves a number of word-formation patterns with similar or identical meaning.

---

[1] However, approximation in general is not necessarily derived from similative items such as in the cases at hand. Other sources of approximative morphological markers include diminutives and degree markers as well as quantitative items, for instance.

2. Basic theoretical assumptions

Approximative, comparative, and evaluative morphological expressions share the basic structure on which their specific meanings are built: They are binary, with one constituent naming the basic concept to be modified and the other realizing the meaning change in question.

Evaluation is used to express speaker attitude in terms of the degree to which a concept is expressed or evaluated in relation to a standard or default value (e.g., Grandi & Körtvélyessy 2015). This includes more quantitative and objectively traceable evaluations as well as qualitative and subjective ones. Typical manifestations of evaluation are diminution and augmentation, i.e., the assessment of a concept as smaller/lower or larger/higher than the respective default value. Pejoration and melioration, in turn, serve the speaker's pejorative (negative) or valorative (positive) evaluation. However, quantitative and qualitative perspectives cannot always be separated: Diminution, for example, as belittling, is often associated with either a positive or negative evaluation. In this respect, evaluative morphology is often also expressive (cf., e.g., Scherer 2019), in the sense of Foolen (1997: 15): "It is the emotional feeling of the speaker that is expressed and communicated in the expressive function."

In West Germanic, the most important morphological means of evaluation are derivation and compounding whereas other languages also exploit processes such as reduplication or inflection (cf. Grandi & Körtvélyessy 2015, for instance). Scalise (1984) presented some assumptions about the formal properties of evaluative morphology. He claimed that evaluative processes never change the syntactic category of the base word. This assumption of categorial neutrality can be explained by the idea that evaluative markers have a modifying function rather than that of creating new lexemes (cf., e.g., Grandi 2015: 76). At the same time, however, there has been counter-evidence to this assumption in the literature for a long time since it is well known that there are also occasional evaluative processes with a change of word class (cf. Bauer 1997; Grandi 2015 and 2017, for discussion). Categorial neutrality, however, might well be a feature of approximative morphological processes. If true, this could possibly provide a means to differentiate approximation from evaluation.

Comparison is one of the most important functional-semantic classes of adjective formation (for German word-formation see Fleischer & Barz 2012; Kühnhold et al. 1978, for instance). Denominal comparative adjectives make a nominal concept available for comparison with another nominal concept and express similarity with the base noun. The similarity can either refer to individual salient properties or to the category as a whole (cf. Bauer et al. 2013: 312), as in the case of *childish* (with respect to unconcern or petulance) vs. *childlike* (with respect to the child in general). The similarity may vary in scope (Kühnhold et al. 1978: 321–354): (a) Reference identity of the two categories (*eine katastrophale Niederlage* 'a catastrophic defeat': i.e., something being both a catastrophe and a defeat at the same time; *ein flegelhafter Junge* 'a boorish boy': a person which is both a boy and a boor, (b) Similarity with respect to individual salient features or the entire category without reference identity (*das balladeske Lied* 'the balladesque song': a song which is not a ballad but shows some features typical of ballads; *der grippale Infekt* 'the flu-like infection: an infection that is not the flu but flu-like'; *die gummiartige Masse* 'the rubbery mass', *das sackförmige Kleid* 'the baggy dress', *das nixenhafte Geschöpf* 'the mermaid-like creature', *die lederne Haut* 'the leathery skin', *der polizeimäßige Hausmeister* 'the police-like janitor', *ein sommerlicher Herbst* 'a summery autumn') or (c) Standard norms and values of the base noun are met (*die sommerliche Bekleidung* 'the summery dress', *die planmäßige Abfahrt* 'the scheduled departure', *das naturhafte Leben* 'the nature-like life', *die polizeimäßige Verhandlung* 'the police-like trial').

These three groups, especially the first two, are largely realized by the same suffixes. In addition, the characteristics of group (b) are found in many German comparative compounds with heads such as *-ähnlich* ('-like'), *-getreu* ('true to'), *-typisch* ('typical') and *-gleich* ('-like'), e.g., *grippeähnlich*, *engelsgleich* ('flu-like, angel-like').[2] In all of these cases, the right constituent determines the word class of the complex word. These comparative formations make a nominal concept available in the form of an adjective.

The meaning or function of **approximation** has been described as 'resemblance', 'similarity', 'vagueness' (also intentional), 'imitation and fakeness' as well as 'attenuation' (as

---

[2] In addition, there are numerous comparative formations with qualitative adjective heads in adjectival compounds, which are not relevant in our context, since here the comparison is made with respect to the adjectival head, e.g., *lammfromm* ('pious as a lamb'), *grasgrün* ('grass-green'), *aalglatt* (lit. eel smooth, 'slippery as an eel').

reduced degree of a quality). So, approximation expresses a comparison and similarity to a certain concept with respect to one or more properties. It is essential, however, that approximative formations always express that category X – however great the similarity – is ultimately not present, which is why approximation is also referred to as privative. For example, Eitelmann et al. (2020: 805) emphasize that "in such cases, *-ish* does not serve to denote an unequivocal relatedness as with the associative sense, but on the contrary an ultimate dissimilarity". Similarly, the meaning of the approximative marker *fake-* has been described as "(a) FAKE X is not (an) X" (Cappelle et al. 2018: 9; cf. also Van Goethem & Norde 2020). So *Scheinehe* ('sham marriage') or *eheähnliche Verbindung* ('marriage-like union') do exactly not denote a marriage (even if many features of a marriage are present), a *pseudo-contract* or *fake-contract* is not a contract, and *virus-like particles* are not viruses in the strict sense. For this reason, constructions like *ein virusähnliches Virus* ('a virus-like virus') or *Er war einer grippeähnlichen Grippe erkrankt* ('He was sick of a flu-like flu') are strongly marked, if not semantically ill-formed.

As mentioned above, in the literature, approximation is partly considered as an independent category, but partly also as a variant of evaluative morphology (cf., e.g., Grandi & Körtvélyessy 2015). In Kühnhold et al. (1978: 188ff), on the other hand, such cases are classified as privative word-formation alongside other meaning groups such as comparison or gradation. Thus, relevant examples can be found there in the section on 'negation and restricted negation' under the headings 'is not, but appears to be' (e.g., *pseudo-authentisch* 'pseudo-authentic', *scheinfromm* 'sanctimonious'), 'only halfway, barely, almost' (*semi-professionell* 'semi-professional', *halblaut* 'semi-loud'), and 'not but nearly/approximately/almost' (*quasi-stationär* 'quasi-stationary', *paramilitärisch* 'paramilitary').

The formal realization of approximation is heterogeneous. A large part of the relevant examples from the literature are prefixes, prefixoids, confixes and free stems in non-head position, for instance German *schein-* ('appear'), *halb-* ('half'), Dutch *nep-* ('sham, fake'), *namaak-* ('imitate'), *kunst-* ('artificial'), *imitatie-* ('imitation'), Italian *simil-* ('similar to') as well as loan elements that are used in various languages such as *semi-*, *para-*, *pseudo-*, *fake-*, *quasi-*, *sub-*, e.g., Van Goethem & Norde (2020); Masini & Micheli (2020); Cappelle et al. (2023). As left constituents they are not category-changing and result in modification processes. In addition, there are also some suffixal approximative patterns, among which the

English suffix *-ish* which has – besides other meanings – an approximative reading (*freeish*, *greenish*), cf. Eitelmann et al. (2020). The authors point out that the approximative reading of *-ish* is found mainly with adjectival bases, so that there is no word class change here either. At the same time they also give examples of approximative *ish*-derivatives on the basis of numerals or nouns (*fourteenish*, *1984-ish*), i.e., with word class changes. Finally, the Italian reduplicative construction N-*non*-N (Masini & Di Donato 2023) is phrasal in nature and it does not imply a change of the syntactic category.

3. Hypotheses on approximation and comparison

While evaluative morphology (as a functional domain) as well as comparison (as a semantic class) are established and clearly distinguishable due to their semantic-functional and morphological properties, this is less clear in the case of approximation. Thus the question arises whether approximation is actually a category in its own right or rather a variant of one of the other two categories.

Besides the question to what extent not changing the word class is constitutive for approximative morphological processes (in other words, whether only modification processes can be approximative), the meaning of the approximative markers themselves must also be clarified. The known examples are stems or prefixoids such as *schein-*, *quasi-*, *nep-*, or *kunst-*, confixes such as *pseudo-* or *semi-*, or prefixes such as *sub-* or *para-* (note that these assignments are neither unambiguous nor uniform in the literature). All of these markers have a lexical meaning. This distinguishes them from prototypical derivational suffixes that do not have a lexical meaning, but only or primarily a grammatical function, sometimes also called a 'grammatical meaning'.[3] The semantics of formations with approximative left-hand constituents thus results compositionally from the lexical meaning of the constituents (something similar can be assumed for the N-*non*-N construction in Italian). The approximative uses of the English suffix *-ish* form (again) an exception to these observations.

---

[3] Both the demarcation between derivational affixes and stems and between lexical and grammatical meaning is far from clear and has been the subject of extensive discussion in the literature, in particular regarding the notion of affixoid (see for example Hüning & Booij 2014). However, this point cannot be developed further here.

Moreover, approximative markers such as *fake-*, Dutch *nep-*, or German *schein-* (all meaning 'fake', 'sham') do not only express the meaning 'is not, but appears to be' (cf. Section 2) but often also seem to convey the intentional pretense of the existence of a category, in particular with the intention to fool somebody (for a more detailed discussion see Van Goethem & Norde 2020). We will discuss whether this is a semantic feature of approximative meaning in general.

The guiding questions mentioned above will be examined in the following using data from the German-Dutch-English language comparison. We base our study on the following hypotheses:

1. Approximativity is derived from similarity.

2. Approximative patterns may be evaluative, comparative patterns are not.

3. Approximation in the narrower sense is characterized by intentionality: Approximative formations explicitly reject to be X.

4. There are markers that are exclusively approximative or comparative, as well as those that can be both, depending on the context.

5. Approximation is also possible for word-formation patterns with word class changes.

6. There are also non-lexical approximative markers.

We discuss these questions for the Dutch suffix *-achtig* and competing forms in Dutch, as well as for the equivalents in German and English. The starting point is the observation that in connection with nominal bases, *-achtig* sometimes allows approximative interpretations, e.g., *virusachtige deeltjes* ('virus-like particles'), *migraine-achtige hoofdpijn* ('migraine-like headache'). Therefore, besides English *-ish*, *-achtig* might be another approximative suffix although its lexical meaning has been lost in the course of history and although it can cause word class change.

4. Pilot study on approximative adjective formation in Dutch and German

The starting point of our pilot study on complex adjectives with an approximative meaning is the Dutch suffix *-achtig*. Adjectives like *virusachtig* or *migraine-achtig* correspond to adjectives with *-like* in English: *virus-like particles*, *migraine-like headache*.

As pointed out in Hüning (2004), the Dutch suffix *-achtig* can be used to form adjectives on the basis of nouns, adjectives and verbs, and the resulting adjectives can be grouped into several semantic categories. In this paper we focus on one group, namely denominal adjectives with *-achtig* with a comparative or approximative meaning.

As we are especially interested in the equivalence of this pattern with word-formation patterns in German, we first looked for translational equivalents. Candidates for corresponding suffixes are the etymologically related German suffixes *-haft* and *-artig* (cf. Pijnenburg 1993; Maesfranckx & Taeldeman 1998; Hüning 2004). They are both used in comparative contexts: *griepachtige symptomen – grippeartige Symptome* ('flu-like symptoms'), *lenteachtig weer – frühlingshaftes Wetter* ('spring-like weather'). However, in contexts for which we are inclined to assume an approximative interpretation for the adjective, *-achtig* often seemed to correspond to other German word-formation patterns, especially compounding with *-ähnlich*. *Ähnlich* is an adjective meaning 'similar, like X (with regard to certain features)', e.g.: *Er ist seinem Vater sehr ähnlich* 'he is very similar to his father; he resembles his father very much'. For this reason, compounding with *-ähnlich* might be another case of a comparative/approximative word-formation pattern.

As mentioned above, Dutch *leerachtig* 'leatherlike' can not only be used comparatively but also approximatively (cf. examples (1)–(3)). This approximative use is also found with German *lederähnlich*, cf. (4)–(5):

(4)   Das Armaturenbrett ist künftig mit einem **lederähnlichen** Material mit Doppelnähten eingefasst. (deTenTen18)

'In the future, the dashboard will be trimmed with a leather-like material with double stitching.'

(5)   Unser Familienunternehmen bleibt seiner Tradition treu und konzentriert sich ausschließlich auf die Veredelung von Leder und **lederähnlichen** Materialien zu hochwertigen Armbändern für Uhren. (deTenTen13)

'Our family business remains true to its tradition and focuses exclusively on the finishing of leather and leather-like materials into high-quality bracelets for watches.'

Lederähnlich here means that the material is 'very much like leather, almost leather' (with the implication: but it is not), which makes it a good example of a denominal adjective with approximative semantics.

4.1 Methods and data

We tried to compare the relevant markers in the different languages by using parallel corpora on sketchengine.eu, but since we are dealing with low-frequency phenomena, they turned out to be too small and their focus on administrative and technical texts proved problematic for our queries.

Therefore, we used the much bigger and more diverse 'TenTen Corpora' (Jakubíček et al. 2013, available from sketchengine.eu). First, we compiled a list of English adjectives with *-like*. This word-formation element has been described in handbooks as a means for expressing approximative semantics in the adjectival domain. When attached to adjectives, *-like* seems "to mean not so much 'similar to X' but 'approximating X'" (Bauer et al. 2013: 313). We assume that this 'approximating X' reading is also present in denominal formations and that *-like* is an example of a category changing marker with approximative semantics, cf. (6).

(6)   Tamiflu used to treat **flu-like** illnesses caused by agents other than influenza viruses Types A and B virus in humans. (enTenTen20)

This list of *-like*-adjectives served as a basis for a selection of base words from different categories. We looked for the translational equivalents of these words combined with an adjectival suffix in the German and Dutch corpora. This way we collected complex adjectives corresponding to X-*like* in the three languages.

We chose 100 base words from different semantic categories, in particular:

– diseases (*aids*, *asthma*, *flu*, *pneumonia*, *stroke*)
– mass nouns (*honey*, *marble*, *meat*, *milk*, *velvet*)
– animal names (*bird*, *eagle*, *panda*, *peacock*, *wolf*)
– names for persons, functions etc. (*boss*, *dandy*, *professor*, *teacher*, *uncle*)

In order to distinguish comparative from approximative morphology, we were especially interested in competing word-formation processes. For our queries we used the lemma search offered by the TenTen Corpora and we looked for adjectives ending in:

– English: *-like, -esque, -ish*
– Dutch: *-achtig, -esk, -erig, -like, -matig*
– German: *-artig, -ähnlich, -esk, -haft, -like, -mäßig*

We hypothesized that -*like* might be used as a loan suffix in German and Dutch, and therefore included it in our queries in the respective corpora. In English, this suffix does not always express similarity. *Warlike*, for example, translates as both *kriegsähnlich* ('similar to war)' and *kriegerisch* ('belonging to war'), as in *warlike violence* ('kriegerische Gewalt'). We will not discuss the latter use.

Very frequent suffixes like English -*y*, Dutch -*ig* or German -*lich* and -*ig* were not systematically taken into account. They are, however, semantically extremely versatile and therefore show some overlap with the categories we are mainly interested in.

We subsumed the inflected forms under the citation form of the adjective. Differences in spelling were not taken into account. Therefore forms with and without hyphens are listed together (*honey-like*, *honeylike*); the more frequent forms are used as citation forms.

We compiled files with KWIC-concordances for the three languages, which served as a basis for frequency lists of the different forms and for our further considerations. We used the following corpora on sketchengine.eu:

Tab. 1: Corpora used for this study

corpus (sketchengine.eu)	size (in million words)
English Web 2020 (*enTenTen20*)	38,149
German Web 2013 (*deTenTen13*)	16,526
Dutch Web 2014 (*nlTenTen14*)	2,254

Note the enormous differences in size of the corpora; this might influence the results.

As mentioned above, we started with a list of English denominal adjectives with -*like*. In the tables below, we present some frequency data to illustrate the distribution of the different markers in English, German and Dutch for our sample.

Tab. 2: Absolute frequencies of English denominal adjectives in *enTenTen20*

ENGLISH (enTenTen20)	*-like*	*-ish*	*-esque*
AIDS	418	–	–
flu	9775	101	–
parkinson	243	–	2
stroke	899	3	–
honey	1025	78	2

ENGLISH (enTenTen20)	*-like*	*-ish*	*-esque*
marble	817	4	3
milk	414	32	–
velvet	724	14	–
wood	1205	27	–
bird	8476	52	1
eagle	440	14	–
panda	73	4	2
peacock	118	35	4
wolf	2962	3808	1
dandy	54	473	31
professor	57	15	6
student	103	26	9
teacher	132	25	10
uncle	33	9	–

Tab. 3: Absolute frequencies of German denominal adjectives in *deTenTen13*

GERMAN (deTenTen13)	*-ähnlich*	*-artig*	*-esk*	*-haft*	*-like*	*-mäßig*
AIDS ('AIDS')	85	5	–	–	–	–
grippe ('flu')	4369	713	–	1	–	38
parkinson ('Parkinson')	196	36	1	1	–	–
schlaganfall ('stroke')	95	16	–	–	–	–
honig ('honey')	144	716		6	–	4
marmor ('marble')	196	166	1	20	–	1
milch ('milk')	133	173			–	5
samt ('velvet')	106	1524		14	–	–
holz ('wood')	325	731	–	1	–	25
vogel ('bird')	574	449	–	55	–	13
adler ('eagle')	37	52	–	35	–	6
panda ('pande')	9	5	–	2	–	5
pfau(en) ('peacock')	15	35	–	80	–	5
wolf(s) ('wolf')	599	202	–	7	–	12
dandy ('dandy')	1	1	82	583	25	7
professor(en) ('professor')	1	–	–	21	1	5
student(en) ('student')	10	1	–	12	4	18

GERMAN (deTenTen13)	*-ähnlich*	*-artig*	*-esk*	*-haft*	*-like*	*-mäßig*
lehrer ('teacher')	7	6	–	514	3	38
onkel ('uncle')	2	–	1	270	–	7

Tab. 4: Absolute frequencies of Dutch denominal adjectives in *nlTenTen14*

DUTCH (nlTenTen14)	*-achtig*	*-esk*	*-erig*	*-like*	*-matig*
AIDS ('AIDS')	3	–	–	–	–
griep ('flu')	732	–	1061	–	–
parkinson ('Parkinson')	37	–	–	–	–
beroerte ('stroke')	2	–	–	–	–
honing ('honey')	207	–	–	–	–
marmer ('marble')	82	–	–	–	–
melk ('milk')	755	–	13	–	–
fluweel ('velvet')	596	–	–	–	–
hout ('wood')	1368	–	1222	–	–
vogel ('bird')	87	–	–	–	–
arend ('eagle')	8	–	–	–	–
panda ('panda')	2	–	–	–	–
pauw ('peacock')	6	–	–	–	–
wolf(s)|wolven ('wolf')	79	–	–	–	–
dandy ('dandy')	31	10	–	–	–
professor(en) ('professor')	3	–	–	–	–
student(en) ('student')	9	–	1	–	–
leraar ('teacher')	1	–	1	–	–
oom ('uncle')	2	–	–	–	–

4.2 General description of the data

Some things become directly clear when looking at these data sets. First of all, we can conclude that there is some limited competition in English between the adjectival markers *-like* and *-ish* (*flu-like* vs. *fluish*, cf. Bauer, Lieber & Plag 2013: 312–313) . This seems especially true for certain groups of adjectives, since *-ish* is generally rather marginal in the

denominal domain.[4] The use of *-esque* and its equivalent *-esk* in German and Dutch is even rarer.

Comparing the word-formation elements in the three languages reveals a direct equivalence between English *-like* and Dutch *-achtig*. All *-like*-adjectives correspond to an adjective with the suffix *-achtig* in Dutch and *-achtig* seems to be the only suffix used regularly in these contexts.[5] German, on the other hand, has several elements that can be combined with the corresponding base nouns. The elements *-artig* and *-ähnlich* seem to compete in the formation of these adjectives, and *-haft-* and *-mäßig* also occur frequently.

The suffix *-mäßig*, however, mainly occurs in adjectives that are used adverbially, meaning 'with regard to X'[6], cf. (7). Occasionally, it also expresses similarity, cf. (8):

(7)   Mich hat es nämlich **grippemäßig** voll erwischt. (deTenTen13)
        ('I have been hit by the flu.')

(8)   … weil mich so ein doofer **grippemäßiger** Infekt umklammert hat. (deTenTen13)
        ('…because I had such a stupid flu-like infection.')

These German adjectives denote a comparison on the basis of typical features of the noun they are derived from. *-matig*, the Dutch equivalent of *-mäßig*, is – contrary to our expectations – apparently never used in the contexts we are interested in. We did not find any approximative uses of the [N+*mäßig*]$_A$ pattern.

The German suffix *-haft* is used especially with personal and function names (cf. Hüning 2004). This suffix is almost absent in the disease group and marginal with mass nouns. Animacy seems to play an important role for the distribution of *-haft*: diseases < mass nouns < animals < humans. Adjectives with *-haft* usually express similarity with respect to typical or standard norms and values or to salient features of nominal base concept, like in (9):

(9)   Umberto Ecos in einer stilvollen, selten **professorenhaften** Sprache geschriebene
        Essays (deTenTen13)

        'Umberto Eco's essays written in a stylish, rarely professorial language'

---

[4] For *-ish* see the contribution by Eitelmann & Haumann (2023).

[5] There are some lexicalized examples with *-erig* in Dutch which we will not discuss in this paper. *Grieperig* ('fluish') is often used predicatively (*ik ben wat grieperig* 'I've got a touch of flu') and *houterig* ('wooden') is often used metaphorically (*een houterig mens* 'a stiff person').

[6] Adverbial German *-mäßig* and Dutch *-matig* are described in Diepeveen (2012).

There are, however, no approximative uses of [N+*haft*]$_A$ in our sample.

Our hypothesis that *-like* might be used productively as a loan suffix in the other two languages could not be confirmed. We only found very few examples for German, especially with derivatives from personal or function names. They have a comparative meaning.

(10) Zum Interview erscheint Matthias Haller **Professoren-like** mit Anzug und Krawatte – aber unstandesgemäss auf einem Velo. (deTenTen13)

   'Matthias Haller appears for the interview professor-like in a suit and tie – unseemly on a bicycle.'

(11) Sie ist ganz **studentenlike** mit der Mitfahrzentrale unterwegs. (deTenTen13)

   'She is travelling all student-like with the ride-sharing service.'

For Dutch, the yield is even lower (which might be due to the size of the Dutch corpus). In the following example, the use of the adjective could be analyzed as approximative, but it is unclear to what extent this needs to be explained by the form correspondence with the English adjective.

(12) Mijn laatste ontwerp, de zwarte fascinator is van zwart **rubber-like** materiaal. Het gevoel en beweging van rubber ligt heel dicht bij vilt waar veel hoeden van gemaakt worden. (nlTenTen14)

   'My latest design, the black fascinator, is made of black rubber-like material. The feel and movement of rubber is very close to felt which many hats are made of.'

In any case, the general conclusion is that – contrary to our assumption – the possibility of forming *-like* adjectives is not popular in (written) German and Dutch.

This leaves us with the comparison of English *-like* and Dutch *-achtig* on the one hand and with German *-artig* and *-ähnlich* on the other. These are the elements that are used as markers both for comparison and approximation.

4.3 Approximative morphology?

We start with bases denoting a disease. While the adjective *flu-like* has a competitor in *fluish*, *AIDS* is only combined with *-like*. They all have a comparative meaning, but in certain contexts, they also have an approximative reading:

(13)   Mysterious **AIDS-like** illnesses also occurred in primate laboratories a few years before AIDS. (enTenTen20)

(14)   Researchers were able to confirm the **AIDS-like** disease is not contagious. (enTenTen20)

In line with Bauer et al. (2013: 312), contexts like "AIDS-like symptoms" where -*like* indicates similarity to individual salient qualities can be distinguished from cases such as "an AIDS-like disease", where -*like* indicates the similarity to a whole, implying 'but it's not the same'. Therefore, we are inclined to regard this use of -*like* as approximation.

Fluish, on the other hand, is mainly used to indicate 'attenuation' (in the sense of reduction or a reduced degree of a quality). This fits the definition of *fluish* in the Merriam-Webster dictionary[7]: "mildly affected with influenza".

(15)   Came home and ten days later came down with a strange **fluish** ailment and the first fever in years. (enTenTen20)

At the same time, there are also approximative usages, where some kind of holistic similarity is expressed.

(16)   I had too little methadone; after two-and-a-half weeks I felt **fluish**, with cold shivers. (enTenTen20)

Dutch *griepachtig* is also used to express similarity:

(17)   [Frühsommer-Meningoenzephalitis (FSME)] is een ziekte die kan variëren van een **griepachtig** beeld tot een ernstige hersenontsteking. (nlTenTen14)

'TBE is a disease that can range from a flu-like appearance to severe meningitis.'

Griepachtige klachten or *griepachtige symptomen*/*verschijnselen* (all meaning 'flu-like symptoms') refer to aspects of an illness that are similar to symptoms of flu. A more holistic comparison is also possible, resulting in an approximative reading:

(18)   Aangewakkerd door uitbarstingen van **griepachtige** ziekten als de luchtweginfectie SARS en de vogelgriep, roepen vele deskundigen al jarenlang om het hardst dat we ons op het ergste moeten voorbereiden. (nlTenTen14)

---

[7] https://www.merriam-webster.com/ (accessed 14 March 2023).

'Encouraged by outbreaks of flu-like illnesses such as the respiratory infection SARS and avian flu, many experts have been shouting out loud for years that we should prepare for the worst.'

Such uses of *-like* and *-achtig* correspond very well to *-artig* in German. In addition, German also uses adjectival compounds with the head constituent *-ähnlich* in order to express comparison and approximation. Both *-artig* and *-ähnlich* have a comparative meaning. The literal meaning of *-artig* is 'belonging to a species ('Art')'. More generally, however, the meaning of *-artig* can be described as 'similar, like X, showing resemblance in certain qualities or characteristics'. The meaning has thus been bleached and *-artig* has been grammaticalized as a general similarity marker. The grammatical status of *-artig* is, however, not fully clear. It is usually described as a suffix or as a suffix-like [sic!] bound element (cf. Fleischer & Barz 2012: 304). In contrast, *-ähnlich* has a clear lexical comparative meaning ('similar, like X, showing resemblance in certain qualities or characteristics') which is found both when used as an adjective and as compound head. Example (19) illustrates this use:

(19)   Die FSME äußert sich zunächst durch **grippeähnliche** Symptome wie Gliederschmerzen und Fieber. (deTenTen13)

         'TBE is initially manifested by flu-like symptoms such as aching limbs and fever.'

Examples like (20) and (21) show that in addition an approximative reading of *-ähnlich* is also possible.

(20)   Die chronische Uranvergiftung führt zu einem **AIDS-ähnlichen** Immundefekt oder zu Krebserkrankungen, insbesondere Leukämie. (deTenTen13)

         'Chronic uranium poisoning leads to AIDS-like immunodeficiency or cancer, especially leukemia.'

(21)   Die ersten SIV-Isolate wurden bei Rhesusaffen gefunden, die in Gefangenschaft in Zoos oder Primatenzentren gehalten wurden und an **AIDS-ähnlichen** Krankheiten verstorben waren. (deTenTen13)

         'The first SIV isolates were found in rhesus monkeys kept in captivity in zoos or primate centers that had died from AIDS-like diseases.'

An approximative interpretation of *-ähnlich* can also be found in formations with mass nouns, cf. (22)–(24).

(22) In France, together with orange, lemon and sugar a **honey-like** jelly is made from the flowers and used as a spread. (enTenTen20)

(23) Eine **honigähnliche** Zuckerpaste wird gegen die natürliche Wuchsrichtung des Haares aufgetragen und dringt so bis zum Haarschaft ein. (deTenTen13)

'A honey-like sugar paste is applied against the natural direction of hair growth, penetrating all the way to the hair shaft.'

(24) De suikers en het vruchtvlees concentreren, waardoor er bij het persen een bijna **honingachtig** vocht ontstaat. (nlTenTen14)

'The sugars and pulp concentrate, producing an almost honey-like liquid when pressed.'

Honey-like, *honigähnlich* and *honingachtig* are used attributively with a noun X and compare X with honey. The implication is: X is almost honey, but ultimately it is not. In (24), this approximative reading is strengthened by the preceding adverb *bijna* ('almost').

In German, *honigartig* can be used in the same way, although *honigähnlich* seems to be preferred, since it expresses the approximative meaning more clearly ('almost honey, but not really'). *Honigartig* as used in (25), on the other hand, leaves open the question whether X actually belongs to the category 'honey' or not.

(25) Wer es süß mag, kann die frischen Blüten zu einem Gelee oder einem **honigartigen** Sirup für einen Brotaufstrich verarbeiten. (deTenTen13)
'For those who like it sweet, the fresh flowers can be made into a jelly or a honey-like syrup for a spread.'

In other contexts, the noun that is modified refers to a property that is part of something else: X is similar to honey with respect to some property Y. Property Y is denoted by the noun which in turn is modified by the adjective. In the following examples this property is the color and the *honey-like color* refers to rosins, earwax and pale ale.

(26) Many brands sell rosins in two colors, at the same price: a light, **honey-like** color, and a darker color, almost like licorice. (enTenTen20)

(27) Normaal heeft het oorsmeer een **honingachtige** kleur. (nlTenTen14)

'Normally, earwax has a honey-like color.'

(28) Im Glas brilliert das East India Pale Ale durch seine appetitliche **honigartige** Farbe. (deTenTen13)

'In the glass, the East India Pale Ale shines with its appetizing honey-like color.'

This means that the modified noun (the color) functions as a 'tertium comparationis': rosins, earwax and pale ale are similar to honey with respect to the color. In such contexts, the adjective is never approximative since rosins, earwax and pale ale are never 'almost honey'. Thus, if an indirect comparison with individual properties introduced via a 'tertium comparationes' is involved rather than a direct comparison of whole categories, an approximate interpretation never emerges. Indirect comparisons referring to individual properties are thus clear examples of comparative meanings that cannot receive an approximative interpretation. German typically uses *-artig* in these comparative constructions, but *-ähnlich* is also possible:

(29) Der Frühlings-Darjeeling Tee hat eine hellgelbe bis **honigähnliche** Farbe und ein lieblich-blumiges Aroma mit einem Hauch Muskat. (deTenTen13)

      'Spring Darjeeling tea has a light yellow to honey-like color and a lovely floral aroma with a hint of nutmeg.'

However, because of its lexical meaning, German *-ähnlich* often seems to be more appropriate for the expression of approximation. Here are some examples with mass nouns as the base of the adjective:

(30) Moleskin ist ein **samtähnlicher** Stoff, jedoch mit geringerer Florhöhe als Samt. (deTenTen13)

      'Moleskin is a velvet-like fabric, but with a lower pile height than velvet.'

(31) Die Insel Brač ist bekannt für ihren weißen, **marmorähnlichen** Kalkstein (deTenTen13)

      'The island of Brač is known for its white marble-like limestone'

(32) Angeboten wird Cordon Bleu, Roulade und Truthahn – hergestellt aus Weizen-Eiweiß und Soja, die zusammen eine **fleischähnliche** Substanz ergeben. (deTenTen13)

      'Cordon bleu, roulade and turkey are offered - made from wheat protein and soy, which together create a meat-like substance.'

For the adjective in the last example, *fleischähnlich*, the approximative meaning has become almost lexicalized in times of veganism. Its counterpart *fleischartig* is sometimes used for a very different meaning:

(33) Möglicherweise hat auch die damals im AStA-Kühlschrank gefundene **fleischartige** Substanz den Gammelfleischskandal ausgelöst. (deTenTen13)

'It is also possible that the meat-like substance found in the AStA refrigerator at the time triggered the rotten meat scandal.'

Fleischartig here does not mean 'almost Fleisch'. The word refers to substances that no longer have much to do with meat. We might call this usage 'disproximative' (in the sense of Cappelle et al. 2023).

Besides *fleischähnlich*, there are many more examples of 'real world approximation', i.e. words denoting substitutes for products that people want to avoid. *Kaffeeähnliche oder milchähnliche Produkte* ('coffee-like or milk-like products') belong to a certain lifestyle. Such examples indicate a possible functional division of labor between the two morphological markers according to which *-ähnlich* has (or gets) the function of approximation marker while *-artig* is (or becomes) the default marker for comparison.

To investigate this hypothesis, we took a closer look at a subset of 13 lexemes from our sample. For each lexeme we analyzed the contexts for the derivation with *-artig* and the compound with *-ähnlich*. We labeled them as 'approximative' or 'comparative' use according to the above mentioned criterion (at the beginning of 4.3) based on the argumentation in Bauer, Lieber & Plag (2013: 312). Distinguishing the uses is, however, not always easy and sometimes impossible. We labeled such uses as 'other/uncertain'. For each pair of words, we examined 50 randomly selected examples per word. We also looked at 50 randomly selected examples of their counterparts in Dutch (with *-achtig*) and in English (with *-like*) (cf. Tab. 5–7).

Tab. 5: The meaning of German adjectives with *-ähnlich* and *-artig* (in %, N=50)

	GERMAN *-ähnlich* vs. *-artig*	comparison	approximation	other/uncertain
1	*Asthmaähnlich* ('asthma-like')	0	94	6
	asthmaartig ('astma-like')	4	94	2
2	*butterähnlich* ('butter-like')	46	46	8
	butterartig ('butter-like')	42	46	12
3	*fleischähnlich* ('meat-like')	56	38	6
	fleischartig ('meat-like')	54	34	12
4	*grippeähnlich* ('flu-like')	32	64	4

	GERMAN *-ähnlich* vs. *-artig*	comparison	approximation	other/uncertain
	grippeartig ('flu-like')	18	80	2
5	*gummiähnlich* ('rubber-like')	14	86	0
	gummiartig ('rubber-like')	8	84	8
6	*holzähnlich* ('wood-like')	44	54	2
	holzartig ('wood-like')	10	86	4
7	*honigähnlich* ('honey-like')	40	58	2
	honigartig ('honey-like')	72	18	10
8	*kaffeeähnlich* ('coffee-like')	38	58	4
	kaffeeartig ('coffee-like')	58	32	10
9	*katzenähnlich* ('cat-like')	16	70	14
	katzenartig ('cat-like')	14	76	10
10	*milchähnlich* ('milk-like')	18	82	0
	milchartig ('milk-like')	24	72	4
11	*plastikähnlich* ('plastic-like')	10	84	6
	plastikartig ('plastic-like')	24	56	20
12	*vogelähnlich* ('bird-like')	16	80	4
	vogelartig ('bird-like')	8	84	8
13	*zombieähnlich* ('zombie-like')	4	88	8
	zombieartig ('zombie-like')	10	84	6

Tab. 6: The meaning of adjectives with English *-like* (in %, N=50)

	ENGLISH *-like*	comparison	approximation	other/uncertain
1	*asthmalike*	4	96	0
2	*butterlike*	44	54	2
3	*meatlike*	40	58	2
4	*flulike*	4	94	2
5	*rubberlike*	22	76	2
6	*woodlike*	18	78	4
7	*honeylike*	46	44	10
8	*coffeelike*	40	54	6
9	*catlike*	30	58	12

10	*milklike*	14	82	4
11	*plasticlike*	10	84	6
12	*birdlike*	8	78	14
13	*zombielike*	14	64	22

Tab. 7: The meaning of adjectives with Dutch *-achtig* (in %, N=50)

	DUTCH *-achtig*	comparison	approximation	other/uncertain
1	*asthma-achtig* (N=0) ('asthma-like')	–	–	–
2	*boterachtig* ('butter-like')	42	46	12
3	*vleesachtig* ('meat-like')	38	54	8
4	*griepachtig* ('flu-like')	2	94	4
5	*rubberachtig* ('rubber-like')	32	66	2
6	*houtachtig* ('wood-like')	12	78	10
7	*honingachtig* ('honey-like')	66	28	6
8	*koffieachtig* (N=18) ('coffee-like')	56	39	5
9	*katachtig* ('cat-like')	10	84	6
10	*melkachtig* ('milk-like')	28	62	10
11	*plasticachtig* ('plastic-like')	26	64	10
12	*vogelachtig* ('bird-like')	10	84	6
13	*zombieachtig* ('zombie-like')	14	76	10

These results are somewhat inconclusive. They do not support our hypothesis of a systematic difference between the two word-formation patterns, i.e. *-ähnlich* mainly used for approximation and *-artig* mainly used for comparison. Instead, the distribution seems to obey other criteria and seems to be lexically driven in many cases. Overall, we found (for both patterns) much more approximation than comparison. The predicted distribution could only be proven for *Honig* ('honey') and *Kaffee* ('coffee'), both cases of 'real world approximation'. For instance, while *honigähnlich* ('honeylike') has a preference for the approximative interpretation (58% of the cases), *honigartig* ('honeylike') has this meaning in only 18% of the cases. For *fleischähnlich* ('meatlike'), however, the picture is clearly different, with the approximative reading in only 38% of all cases, which is almost identical to the distribution for *fleischartig* (with 34% approximative uses). Other base nouns form adjectives with a clear preference for the approximative use, like *gummiähnlich* (86%) or *gummiartig* (84%), both meaning *rubberlike* which is also mostly used approximatively in

English. Such cross-linguistic preferences for the approximative use can also be found in the case of *asthmalike* and *asthmaartig*/*asthmaähnlich* and for *birdlike* and *vogelartig*/*vogelähnlich*. Dutch *vogelachtig* joins this picture, but *asthma-achtig* does not occur at all.

We also found cases that ran contrary to our expectations: *woodlike*, *houtachtig* and *holzartig* are used approximatively in about three-fourth of the cases, while *holzähnlich* shows this reading in only half of the cases. *Holzartig* is, however, used in a particular collocation (*holzartige Biomasse* 'the ligneous biomass') in 19 out of 50 examples, which means this is not only a lexical but also a collocational effect (the combination with *Biomasse* does not occur with *holzähnlich*).

An important insight we gain from this little investigation is that the principle of isomorphism, according to which every linguistic form is ideally assigned a unique meaning or function, is at best an ideal, an example of linguistic wishful thinking. In reality, speakers often do not have any problems with synonymous forms. As far as our data allow this conclusion, this does not only concern the co-existence of single lexemes, but (also) of patterns. For the word-formation patterns with *-ähnlich* and *-artig*, there is no indication of the emergence of specific semantic niches, nor do the patterns block each other. Rather, the two patterns co-exist and in many cases show similar preferences of usage, which seems to be lexically determined.

Thus, in our data there are many synonymous uses of adjectives with *-artig* and *-ähnlich*. This can be illustrated by the following examples, where *kaffeeartig* and *kaffeeähnlich* show the same approximative reading:

(34)   Aus den ausgelösten und gerösteten Nüssen kann **ein kaffeeartiges Getränk** bereitet werden. Auch bei den amerikanischen Ureinwohnern waren Bucheckern als Nahrung weit verbreitet. (deTenTen13)

    'A coffee-like beverage can be prepared from the roasted nuts. Beech nuts were also widely used as food by Native Americans.'

(35)   Des Weiteren können die Samen geröstet werden und liefern dann **ein kaffeeähnliches Getränk**. Getrocknete Lupinensamen werden im Handel unter der Bezeichnung Tirmis vertrieben. (deTenTen13)

    'Furthermore, the seeds can be roasted and then provide a coffee-like beverage. Dried lupine seeds are sold commercially under the name Tirmis.'

In the above examples, it is clear that the beverages being discussed are not categorized as coffee. In (36)–(37), on the other hand, the boundary between coffee and non-coffee is blurred. Here, the beverages in question are characterized as not being 'real' coffee anymore because of the additional ingredients and because they are served cold. We analyze this as a 'disproximative' reading of the adjectives in the sense of Cappelle et al. (2023); see also example (33).

(36) Im Café sitzen und einen Frappé (oder ein **anderes kaffeeartiges Getränk**) trinken ist eine Lieblingsbeschäftigung der Griechen.

'Sitting in a café and drinking a frappé (or other coffee-like beverage) is a favorite pastime of Greeks.)'

(37) Starbucks hat seine erste Filiale in Seattle eröffnet (sie gehört zu den wichtigsten Sehenswürdigkeiten der Stadt) und von hier aus die ganze USA, und mittlerweile auch den Rest der Welt, mit Caramel Frappuccino, Iced Vanilla Latte und **anderen Kaffeeähnlichen Getränken** überschwemmt.

'Starbucks opened its first branch in Seattle (it is one of the city's most important landmarks) and from here has flooded the entire U.S., and by now the rest of the world, with Caramel Frappuccino, Iced Vanilla Latte and other coffee-like beverages.'

The disproximative meaning makes these adjectives very suitable for expressive contexts and negative connotations, which becomes particularly clear in example (37). Even in such contexts, however, *-artig* and *-ähnlich* seem interchangeable. So far we have not been able to find a useful criterion for distinguishing them systematically.

5. Discussion

In order to answer our initial questions about the definition and distinction of approximation vs. comparison, we now discuss the hypotheses formulated in Section 3.

1. Approximativity is derived from similarity.

Unsurprisingly, the examples in the previous sections have shown that all approximative readings are at the same time also similative, since approximation to a category presupposes similarity with this category. This connection is particularly clear since the markers at hand are similative items. The similarity may relate either to the entire category or to single salient properties, both with or without referential identity.

2. Approximative patterns may be evaluative, comparative patterns are not.

As outlined in Section 2, the category of evaluation includes both more quantitative and objectively traceable as well as qualitative and subjective evaluations. Our hypothesis concerns the quantitative assessment of a category as very similar to another one. In other words, approximation can be used to express the extent to which one category approaches another, as is especially evident in cases such as *semi-*, *sub-*, *half-* etc. It can be reasonably argued that this meaning is not present in comparative patterns since the degree of similarity does not matter here but only the property to which the comparison refers. This is particularly obvious in comparative structures with a 'tertium comparationis' as in (26)–(29). Thus, our data seem to confirm hypothesis (2) since they do not explicitly denote the degree of similarity between the categories in question but just the fact that they are similar. However, some constructions with *-ähnlich* might have a quantitative evaluative flavor, due to the lexical meaning of this element, especially in contexts that suggest a high degree of similarity, such as in example (30), repeated here as (38):

(38)   Moleskin ist ein **samtähnlicher** Stoff, jedoch mit geringerer Florhöhe als Samt. (de-TenTen13)

'Moleskin is a velvet-like fabric, but with a lower pile height than velvet.'

3. Approximation in the narrower sense is characterized by intentionality: Approximative formations explicitly reject to be X.

The criterion of intentionality and rejection seems to be restricted to approximative markers in the strict sense. Approximative patterns such as *simil-*, *fake-*, *pseudo-*, *quasi-*, *kunst-* etc. explicitly reject to be X. More precisely, the approximative marker (a) invokes the possibility that category identity might be present, while (b) at the same time clearly rejecting it (cf. Section 2). (Obviously, there are differences between the various left markers with respect to the exact nature of this pretending and negating, but this cannot be further addressed here, cf. e.g. Cappelle et al. 2023). Examples such as (39)–(40) show that this is not necessarily the case with comparative patterns: The dilution might be water or just similar to water, the particles might be viruses or just like viruses.

(39)   a.   Insuman Infusat is een heldere en kleurloze oplossing voor injectie, met geen vaste deeltjes zichtbaar en een **waterachtige** consistentie.

     b.   Insuman Infusat ist eine klare, farblose, **wässrige** Injektionslösung ohne sichtbare Teilchen.

        'Insuman Infusate is a clear, colorless, aqueous solution for injection without visible particles'

(40)   a.   Es handelt sich um L1-Protein in Form von **virusähnlichen** Partikeln, ja, möglicherweise sogar um richtige Viren.

     b.   Dit is L1 eiwit in de vorm van **virusachtige** deeltjes, mogelijk zelfs echte virussen.

        'It is L1 protein in the form of virus-like particles, possibly even real viruses.'

Thus, the patterns discussed here do not belong to this group. In fact, an explicit rejection of being X is not found in any of the examples discussed here.

4. There are markers that are exclusively approximative or comparative, as well as those that can be both, depending on the context.

All the comparative markers studied here (-*like*, -*ähnlich*, -*artig* and -*achtig*) also allow approximative readings. Our study suggests that the question whether they are predominantly used comparatively or with an approximative reading mainly depends on the respective base nouns. Thus, they are not exclusively comparative markers. This might, however, be different for other comparative markers such as -*haft* and -*mäßig* that were not in the focus of this study.

Note also that the markers investigated here are all suffixes or other kinds of right constituents. The finding that they are comparative markers which also allow approximative readings parallels English -*ish*, another suffix which has both a comparative and an approximative meaning (cf. Eitelmann et al. 2020: 805).

While neither exclusively approximative nor comparative markers have been attested in our study, we nevertheless assume that there are exclusively approximative markers, i.e. markers that have been referred to as 'approximative markers in the strict sense' above, e.g. *fake-*, *simil-*, *quasi-* etc. Although similarity is an essential part of their meaning, as argued before, they do not allow purely comparative readings.

5. Approximation is also possible for word-formation patterns with word class changes.

-like, *-ähnlich*, *-artig* and *-achtig* are head constituents and may change the word class, as in our data. They also allow approximative readings. Approximative readings are therefore also possible for patterns with a change of word class. At the same time, it has been found that approximative markers in the strict sense, i.e. markers that are exclusively approximative, are all left constituents and thus not word class changing.

6. There are also non-lexical approximative markers.

As argued above, non-lexical markers such as *-achtig* and *-artig* may have approximative readings (just as the English non-lexical marker *-ish*). On the other hand, approximative markers in the strict sense are all lexical markers. In this connection, it is interesting to review German *-artig* and *-ähnlich* and the respective differences since the latter is a lexical marker. Although the results of our pilot study indicate that there is no principled difference between *-artig* and *-ähnlich* with regard to comparative and approximative uses, it seems to us that approximative readings might be more easily available for *-ähnlich* than for *-artig*. This might have to do with the lexical meaning of *-ähnlich*: In addition to the comparative meaning 'showing resemblance in certain qualities or characteristics' (which does not say much about the degree of similarity, cf. hypothesis 2), *-ähnlich* can also be interpreted (also when used as an adjective) as 'nearly but not exactly the same', thus with a high degree of similarity and therefore approximatively.

6. Conclusion

The aim of our study was to contribute to a better understanding of the notion of (morphological) approximation. In particular, our aim was to delineate the notion of approximation from the notion of comparison and similarity. However, we have seen that this is only possible to a limited extent, at least with regard to the data examined here. Similarity, as shown, is a necessary part of the meaning of approximation, and comparative markers can have approximative readings.

In order not to dilute the notion of approximation and diminish its theoretical relevance, it might be better so reserve it for the entities that have been referred to here as "approximative markers in the strict sense" or to distinguish between approximative markers proper and those with an optional approximative readings, as in the present paper.

Approximative markers in the strict sense are left-headed, have lexical meaning, and explicitly express the rejection of category membership.

References

Bauer, Laurie. 1997. Evaluative Morphology: In Search of Universals. *Studies in Language* 21(3). 533–575. DOI: 10.1075/sl.21.3.04bau.

Bauer, Laurie, Rochelle Lieber & Ingo Plag. 2013. *The Oxford Reference Guide to English Morphology* (Oxford Linguistics). Oxford: Oxford University Press.

Cappelle, Bert, Pascal Denis & Manuela Keller. 2018. Facing the facts of *fake*: A distributional semantics and corpus annotation approach. *Yearbook of the German Cognitive Linguistics Association* 6(1). 9–42.

Cappelle, Bert, Robert Daugs & Stefan Hartmann. 2023. The English privative prefixes *near-*, *pseudo-* and *quasi-*: approximation and 'disproximation'. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1). 52–75.

Diepeveen, Janneke. 2012 *Modifying words: Dutch Adverbial Morphology in Contrast*. PhD Dissertation. Berlin: Freie Universität Berlin. https://refubium.fu-berlin.de/handle/fub188/9301 (accessed 14 March 2023).

Eitelmann, Matthias, Kari E. Haugland & Dagmar Haumann. 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.

Fleischer, Wolfgang & Irmhild Barz. 2012. *Wortbildung der deutschen Gegenwartssprache*. 4., völlig neu bearbeitete Auflage. Berlin: De Gruyter. DOI: 10.1515/9783110256659.

Foolen, Ad. 1997. The expressive function of language: Towards a cognitive semantic approach. In Susanne Niemeier & René Dirven (eds.), *The Language of Emotions. Conceptualization, Expression, and Theoretical Foundations*, 15–31. Amsterdam & Philadelphia: Benjamins.

Grandi, Nicola. 2015. The Place of Evaluation within Morphology. In Nicola Grandi & Lívia Körtvélyessy (eds.), *Edinburgh Handbook of Evaluative Morphology*, 74–90. Edinburgh: Edinburgh University Press.

Grandi, Nicola. 2017. Evaluatives in Morphology. In *Oxford Research Encyclopedia of Linguistics*. Oxford University Press. DOI: 10.1093/acrefore/9780199384655.013.250.

Grandi, Nicola & Lívia Körtvélyessy. 2015. Introduction: Why Evaluative Morphology? In Nicola Grandi & Lívia Körtvélyessy (eds.), *Edinburgh Handbook of Evaluative Morphology*, 3–20. Edinburgh: Edinburgh University Press.

Hüning, Matthias. 2004. Concurrentie en equivalentie in de morfologie: het suffix *-achtig*. In Johan De Caluwe, Georges De Schutter, Magda Devos & Jacques van Keymeulen (eds.), *Taeldeman, Man van de Taal, Schatbewaarder van de Taal*, 547–556. Gent: Academia Press.

Hüning, Matthias & Geert Booij. 2014. From compounding to derivation: The emergence of derivational affixes through "constructionalization." *Folia Linguistica* 48(2) (Special Issue "Refining Grammaticalization", ed. by Ferdinand von Mengden & Horst Simon). 579–604. DOI: 10.1515/flin.2014.019.

Jakubíček, Miloš, Adam Kilgarriff, Vojtěch Kovář, Pavel Rychlý & Vít Suchomel. 2013. The TenTen Corpus Family. In *7th International Corpus Linguistics Conference CL*, 125–127.

Kühnhold, Ingeburg, Oskar Putzer & Hans Wellmann. 1978. *Deutsche Wortbildung. Typen und Tendenzen in der Gegenwartssprache. Dritter Hauptteil: Das Adjektiv* (Sprache der Gegenwart 43). Düsseldorf: Pädagogischer Verlag Schwann. DOI: 10.1515/9783110877151.

Maesfranckx, Patricia & Johan Taeldeman. 1998. Polyseem, polyvalent en vaag *-achtig*. In Eric Hoekstra & Caroline Smits (eds.), *Morfologiedagen 1996* (Cahiers van Het Meertens Instituut 10), 84–105. Amsterdam: Meertens Instituut.

Masini, Francesa & Jacopo Di Donato. 2023. Non-prototypicality by (discontinuous) reduplication: the N-non-N construction in Italian. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1). 130–155.

Masini, Francesca & M. Silvia Micheli. 2020. The morphological expression of approximation: the emerging *simil*-construction in Italian. *Word Structure* 13(3). 371–402. DOI: 10.3366/word.2020.0176.

Merriam Webster Dictionary. https://www.merriam-webster.com/ (accessed 14 March 2023).

Pijnenburg, W. J. 1993. Honderd jaar *-achtig*: Over de wording van een samengesteld suffix. *Amsterdamer Beiträge zur älteren Germanistik* 37. 33–48.

Scalise, Sergio. 1984. *Generative Morphology*. Dordrecht: Foris.

Scherer, Carmen. 2019. Expressivität in der Wortbildung. Ein Überblick. In Franz-Josef d'Avis & Rita Finkbeiner (eds.), *Expressivität im Deutschen* (Reihe Germanistische Linguistik 318), 49–74. Berlin & Boston: De Gruyter. DOI: 10.1515/9783110630190-003.

Van Goethem, Kristel & Muriel Norde. 2020. Extravagant "fake" morphemes in Dutch. Morphological productivity, semantic profiles and categorical flexibility. *Corpus Linguistics and Linguistic Theory* 16(3). 425–458. DOI: 10.1515/cllt-2020-0024.

Van Goethem, Kristel, Muriel Norde & Francesca Masini. 2021. The fate of 'pseudo-' words: A contrastive corpus-based analysis. Paper presented at UCCTS 2021 (Using Corpora in Contrastive and Translation Studies), Bertinoro, 9–11 September 2021.

Matthias Hüning
Freie Universität Berlin
Institut für Deutsche und Niederländische Philologie
Habelschwerdter Allee 45
14195 Berlin, Germany
matthias.huening@fu-berlin.de

Barbara Schlücker
Universität Leipzig
Institut für Germanistik
Beethovenstraße 15
04107 Leipzig, Germany
barbara.schluecker@uni-leipzig.de

Francesca Masini & Jacopo Di Donato

Non-prototypicality by (discontinuous) reduplication: The N-*non*-N construction in Italian

Abstract: This paper contributes to the research on the morphological expression of approximation by analysing the discontinuous reduplication pattern N-*non*-N in Italian, giving rise to complex nominals (e.g., *sapone non sapone*, lit. soap NEG soap, 'non-soap/soap-free cleanser'). The analysis is based on a dataset of 4609 tokens and 692 types extracted from corpora and annotated for orthographical, phonological and categorial parameters. Given its unpredictable formal and semantic properties, N-*non*-N is analysed as a semi-schematic and productive construction, which is hypothesized to have emerged microdiachronically from the entrenchment and reanalysis of the expression *tessuto non tessuto* 'nonwoven fabric'. We claim that the N-*non*-N construction bears a general function of 'non-prototypicality' and produces expressions with complex and specific meanings that are generated in context, by deviating from the 'prototypical' N concept in various ways. Despite its productivity, very few N-*non*-N expressions are conventionalised, making this device similar to other approximative strategies and to evaluative morphology in general.

Keywords: evaluative morphology, approximation, non-prototypicality, discontinuous reduplication, negation, Construction Morphology, Italian

1. Introduction[*]

In this paper, we contribute to the research on the morphological expression of approximation by analysing a discontinuous reduplication pattern in Italian that we claim to convey a value of 'non-prototypicality'.

Reduplication is known for having a variety of functions cross-linguistically, including approximation (cf., among others, Moravcsik 1978; Inkelas 2014; Mattiola & Barotto 2023). Some examples (retrieved through the ListTyp database) follow:

---

(1) a. *le'leng* 'black' > *le'leng-le'leng* 'blackish' (Makasar; Jukes 2006: 105)

  b. *ɨŋgwē* 'white' > *ɨŋgwē ɨŋgwē* 'whiteish, rather white' (Ma'di; Blackings & Fabb 2003: 105)

Italian is generally considered as a language with no productive reduplication (cf. WALS: https://wals.info). However, some studies, either implying or advocating for a unified approach to reduplication and repetition, show that reduplication is well attested in this language, performing a variety of functions[1] (for an overview, see Mauri & Masini 2022, who talk about 'replication'). This is true of both 'canonical' and 'non-canonical' reduplication (Stolz 2018), including so-called 'discontinuous reduplication' (Mattiola & Masini 2022), which may be defined as a kind of reduplication "where other morphological material may appear between the reduplicant and the base" (Velupillai 2012: 101).

  The pattern under scrutiny here is precisely a case of 'discontinuous reduplication' displaying a reduplicated noun N with an interposing negation adverb (*non*) that gives rise to a complex nominal:[2]

(2) a. *sapone*   *non*   *sapone*
    soap   NEG   soap
    'non-soap/soap-free cleanser, synthetic detergents (syndets)'

  b. *Per una maggiore azione dermoprotettiva, si consiglia il '**sapone non sapone**' perché privo di sostanze alcaline che possono causare irritazioni e inaridimento cutanei.* [itTenTen16]

    'For a better skin protection, we recommend a **soap-free cleanser**, as it is free of alkaline chemicals, which can cause skin irritation and dryness'

(3) a. *bomboniera*   *non*   *bomboniera*
    party_favour   NEG   party_favour
    'alternative/unconventional party favour'

---

[1] Since we are talking here about 'non-prototypicality', it is worth mentioning that nominal (continuous) reduplication in Italian may convey quite the opposite value, namely prototypicality/authenticity (e.g., *caffè caffè*, lit. coffee coffee, 'genuine coffee') or contrastive focus, as Gomeshi et al. (2004) call it (see Mauri & Masini 2022: 132–134 for a brief discussion and examples).

[2] Italian *non* mainly corresponds to English *not*, but it also appears in formations like *non credente* 'non-believer' or *nonsenso* 'nonsense', where it corresponds to English *non*. For simplicity, we decided to gloss *non* as NEG throughout the paper.

b. *Ci sono però anche delle proposte alternative e tra queste, ad esempio, le "**bomboniere non bomboniere**" costituite da oggettini più particolari e originali come il prodotto biologico, il barattolino di miele, il prodotto tipico, ecc. per chi ama l'anticonformismo "naturale".* [itTenTen16]

'There are, however, alternatives, like, for example, **unconventional party favours**, made of tinier, more unique and original items such as an organic product, a small honey jar, a regional product etc., for those who love a "natural" non-conformism'

The meaning denoted by the N-*non*-N expressions in (2)–(3) can be holistically schematized as 'non-prototypical N'. *Sapone non sapone* in (2) is a cleanser/detergent (thus serving the same function as soap) which is however not a 'proper' soap in that it is not made through saponification (i.e., the combination of natural oils/fats and an alkaline product). *Bomboniere non bomboniere* (3), instead, refers to party favours that depart from canonical objects (typically containing sugared almonds) usually given at weddings and other ceremonies.

We claim that this N-*non*-N pattern – which was identified by Mattiola & Masini (2022) but still lacks a proper description in the literature – has become a specific device used to convey 'non-prototypicality' in Italian.

The paper is organized as follows. Section 2 describes the corpus-based methodology used to study the N-*non*-N pattern, including annotation parameters (orthographical, phonological, categorial). Section 3 illustrates the results in quantitative terms. In Section 4 we discuss the results and we provide a theoretical account in terms of Construction Grammar (Goldberg 1995, 2006) and Construction Morphology (Booij 2010), not without speculating on the why and how this construction emerged. Section 5 contains some concluding remarks.

2. Methodology

In this paper, we offer an in-depth analysis of the N-*non*-N pattern based on naturally occurring data from *CORIS* (http://corpora.dslo.unibo.it/TCORIS) and *itTenTen16* (accessed through the SketchEngine: https://www.sketchengine.eu), two corpora of written contemporary Italian. The data were extracted automatically through a query and then manually revised.

The final dataset contains 4609 tokens and 692 types. The dataset was annotated with the following parameters in order to identify possible constraints and tendencies:

- **Reduplicated form**: the exact reduplicated word form.
- **Reduplicated lemma**: the lemmatized version of the reduplicated form.

ORTHOGRAPHICAL AND PHONOLOGICAL PROPERTIES

- **Hyphen [Orthography]**: the presence ('yes') or absence ('no') of two hyphens separating the three components of the pattern.[3] We also annotated as 'hybrid' the cases that display only one hyphen.
- **Scare quotes [Orthography]**: the presence ('yes') or absence ('no') of scare quotes surrounding the pattern. We did not consider as scare quotes quotation marks signalling true quotations or proper names, such as titles of books, programs, etc.
- **Syllables**: the number of syllables of the reduplicated item.

CATEGORIAL PROPERTIES

- **Category of reduplicated form**: we searched for patterns including items tagged as nouns in the two corpora, and we indeed found mostly nouns ('N'), but a minority of occurrences turned out to be verbs (e.g., *vedo non vedo*, lit. I.see NEG I.see), which were annotated accordingly as 'V' plus TAM information, resulting in two tags ('V_1SG.PRES.IND' and 'V_2SG.PRES.IND'). The output category of the whole construction, instead, is noun for all the examples.[4]
- **Ambiguity**: a number of occurrences turned out to display categorial ambiguity, since in these cases the reduplicated element (the copy) is interpretable not just as a noun but also as a modifier: either an adjective (e.g., *dolce non dolce*, lit. sweet$_N$ NEG sweet$_{N/ADJ}$, where the first *dolce* is a noun but the second *dolce* might be an adjective), a past participle (e.g., *bollito non bollito*, lit. boiled_meat$_N$ NEG boiled_meat$_N$/boiled$_{ADJ}$) or a present participle (e.g., *cantanti non cantanti* lit. singers$_N$ NEG singers$_N$/singing$_{ADJ}$). We marked these cases as

---

[3] The corpora used for data extraction treated hyphens differently in terms of tokenization. We took this difference into consideration and elaborated different queries in order to capture examples with hyphens in both corpora.

[4] An anonymous reviewer raised the question whether a V-*non*-V pattern yielding verbs (instead of nouns) exists in Italian. To the best of our knowledge, the answer is no.

ambiguous ('yes'), whereas those that unequivocally display a noun in both positions were annotated as unambiguous ('no').

– **Category of copy**: when the example is ambiguous (see previous parameter), we annotated the other possible category of the copy, hence 'ADJ' (adjective), 'PART_PAST' (past participle) or 'PART_PRES' (present participle).

3. Results

3.1 Lexical distribution and productivity

As already said, the final dataset consists in 4609 tokens for 692 types. The lexical distribution is uneven since we have a very high number of types with low frequency and very few cases with an extremely high token frequency: *tessuto non tessuto* 'nonwoven fabric' [lit. fabric NEG fabric/woven] occurs 1966 times, amounting to approx. 42% of the total dataset, followed by *vedo non vedo* 'translucent/see-through effect' [lit. see.1SG.PRES NEG see.1SG.PRES] with 508 tokens, and *luogo non luogo* 'non(-anthropological) place' [place NEG place] with 193 tokens. See Fig. 1, displaying a classic Zipfian distribution for types with fq. 10 or higher.
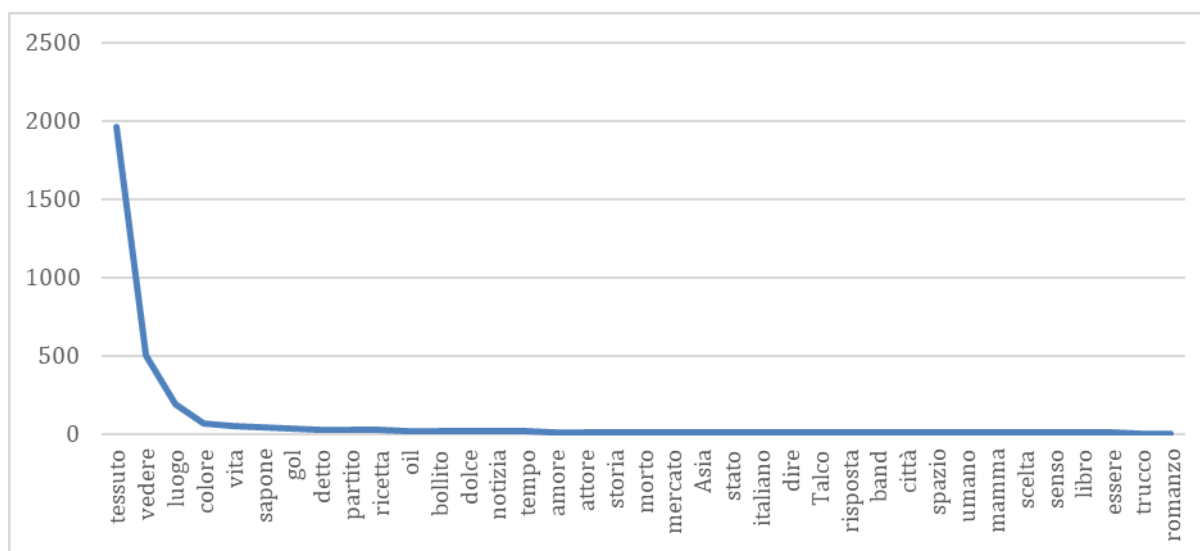
Fig. 1: Lexical distribution of types with fq. 10 or higher

Hapaxes amount to 402 (approx. 58% of the total types), speaking in favour of the productivity of the construction, which is indeed able to generate new expressions.

Productivity measures like type/token ratio (TTR = 0.149) and potential productivity (PP = 0.087)[5] are not very meaningful taken in isolation, but they will be applied to different subsets of our data in Section 4.4. They might also suffer from the unbalance caused by the few very frequent items mentioned above – *tessuto non tessuto* (4a), *vedo non vedo* (4b) and (secondarily) *luogo non luogo* (4c) – which are clearly the 'leaders' in our dataset.

(4) a. *Questo ad evitare, nel tempo, possibili intasamenti. Sopra le pietre va posto uno strato di **tessuto non tessuto**, che ha il compito di impedire il passaggio di terra o di materiale organico, che altrimenti intaserebbe il drenaggio stesso* [itTenTen16]

  'This is to avoid, over time, possible clogging. A layer of **unwoven fabric** should be placed over the stones, which has the task of preventing the passage of soil or organic material, which would otherwise clog the drainage itself'

  b. *I body in pizzo giocano con un intrigante **vedo non vedo**; i più romantici coordinati, arricchiti da raffinati dettagli, sono sempre un must.* [itTenTen16]

  'The lace bodysuits play with an intriguing **see-through effect**; the most romantic coordinates, enriched by refined details, are always a must.'

  c. *La storia è semplice, narrata attraverso un lungo flashback da un 'cantastorie' mentre fa la fila in un "**luogo non luogo**" contemporaneo come l'ufficio postale.* [itTenTen16]

  'The story is simple, told through a long flashback by a 'storyteller' as he lines up in a contemporary "**non(-anthropological) place**" like the post office.'

It is interesting to note that these three expressions are representative of the three types we identified in terms of categories in Section 2: *tessuto non tessuto* is an 'ambiguous' case, since *tessuto* means both 'fabric$_N$' and 'woven' (past participle of *tessere* 'to weave'), hence the whole expression might be interpreted either as a 'true' reduplication mediated by *non*, or as a noun followed by a negated modifier ('unwoven fabric') (cf. Sections 3.3 and 4.4); *vedo non vedo* contains a reduplicated verbal form instead of a noun; and *luogo non luogo* would be an example of the truly reduplicative construction we aim to investigate here, where the second *luogo* is unequivocally a noun.

---

[5] For a discussion and illustration of type/token ratio (TTR) and potential productivity (PP) as productivity scores see Baayen & Lieber (1991) and Baayen (2009).

3.2 Orthographical and phonological properties

As for orthographic properties, we annotated the presence of hyphens within the pattern and scare quotes around it, in order to understand if orthography plays any role in marking certain kinds of expressions.

The vast majority of examples in our dataset does not display hyphens (81%; see Fig. 2) and is not enclosed in scare quotes (88%; see Fig. 3). Hybrid cases are clearly very marginal: in the vast majority of cases, the single hyphen appears before *non*, whereas very few examples display the hyphen after *non* (95 vs. 8 occurrences).

However, the percentage of hyphens and scare quotes increases if we consider only hapaxes, which are more likely to be new formations with a supposedly more creative semantics: approximately half of the hapaxes display at least one hyphen ('yes' or 'hybrid') and approx. 32% of the hapaxes are enclosed in scare quotes.
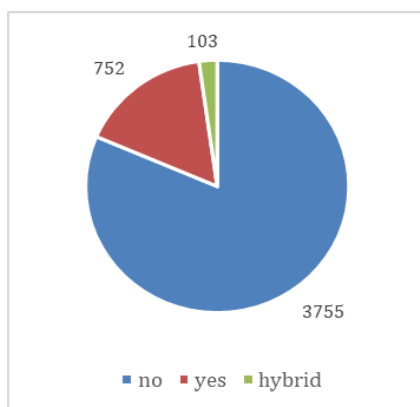
Fig. 2: Presence of hyphen(s) based on tokens
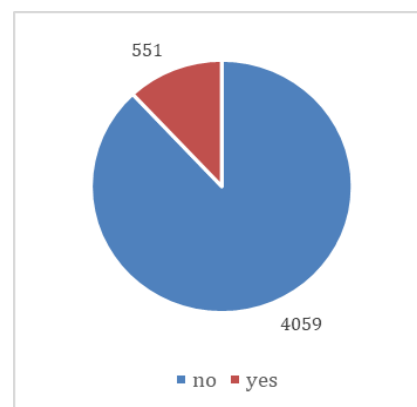
Fig. 3: Presence of scare quotes based on tokens

Note that one and the same item may occur with and without hyphens (cf. (5a) vs (5b)), with and without scare quotes (cf. (6a) vs (6b), where the two instances refer exactly to the same referent).

(5) a. *Poi l'attrice ammette: Sì, potrei lavorare ancora con lui, perché Claudio è un **attore-non-attore**, sovverte continuamente le regole, ha un modo anarchico di intendere il mestiere.* [itTenTen16]

'Then the actress admits: Yes, I could still work with him, because Claudio is an **unconventional actor** [lit. **actor-NEG-actor**], he constantly subverts the rules, he has an anarchic way of understanding the job.'

b. […] *non c'era certo bisogno del film di Antonio Albanese, non avevamo neanche bisogno di* **attori non attori**, *quindi poco professionisti come ormai da troppo tempo si hanno sugli schermi italiani* [itTenTen16]

'[…] there was certainly no need for Antonio Albanese's film, we didn't even need **non-actors** [lit. **actors NEG actors**], namely unprofessional actors as we have had on Italian screens for too long now'

(6) a. *Interverranno lo chef Massimo Bottura, presidente del comitato dei Maestri di Cucina e ideatore di metodi di cottura innovativi, di particolare interesse per il mondo scientifico (famoso il suo* **"bollito non bollito"** *e l'uso in cucina di macchinari per la distillazione sottovuoto per concentrare le essenze)* […]. [itTenTen16]

'Speakers include chef Massimo Bottura, president of the Committee of the Masters of Cuisine and creator of innovative cooking methods, of particular interest to the scientific world (famous is his "**unboiled boiled meat**" [lit. boiled_meat NEG boiled_(meat)] and the use in the kitchen of machinery for vacuum distillation to concentrate the essences) […]' [itTenTen16]

b. *Il* **bollito non bollito** *di Massimo Bottura, ad esempio, pur essendo la quintessenza della semplicità* […] *richiede l'uso di procedimenti non proprio alla portata di tutti.* [itTenTen16]

'Massimo Bottura's **unboiled boiled meat** [lit. boiled_meat NEG boiled_(meat), for example, while being the quintessence of simplicity […] requires the use of procedures not really within everyone's reach.'

Another feature we checked is the number of syllables of the reduplicated form: the syllabic structure may shed light on instantiation preferences, based on the phonological weight of the word to be reduplicated. Our data show that trisyllabic and bisyllabic words are definitely more used in this construction (93% of the tokens, 78% of the types). Much less attested are monosyllabic words (mainly borrowings from English like *gol non gol* or *film non film*) and longer bases of 4 (e.g., *gravidanza-non-gravidanza*, lit. pregnancy NEG pregnancy), 5 (e.g., *appuntamento non appuntamento*, lit. date NEG date) and 6 (e.g., *ideologia-non-ideologia*, lit. ideology NEG ideology) syllables (see Fig. 4).
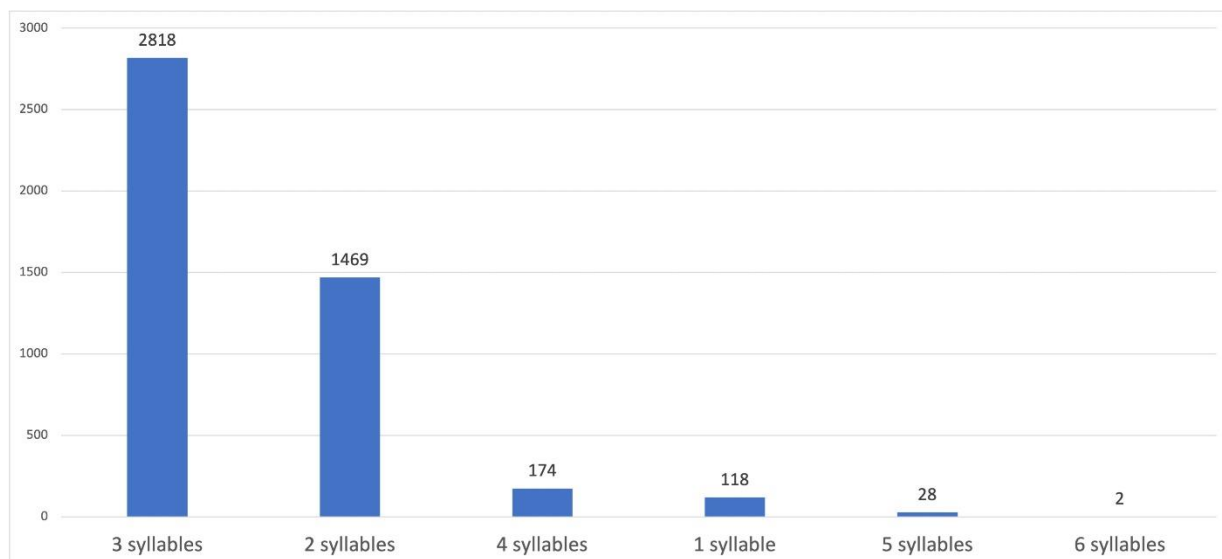
Fig. 4: Number of syllables

The preference for short bases may just reflect a general tendency in the Italian lexicon, so we confronted our data with those that can be retrieved from the NVdB (*Nuovo Vocabolario di Base*)[6]. The sum of trisyllabic and bisyllabic nouns in the NVdB equals 46%. The difference with our dataset is quite marked, so there is some ground to state that the N-*non*-N construction has indeed a preference for bi/trisyllabic, hence 'short' words. This may have something to do with the prosody associated with the construction, an aspect we cannot investigate here given that we are dealing with written languages, but that sounds promising.

3.3 Categorial properties

As anticipated in Section 2, reduplicated items are normally nouns (and noun is also the output category of all our examples), but in some cases (11% of tokens, 1% of types) we retrieved verbal forms (obviously due to wrong POS tagging). In the latter case, basically all examples are 1st person singular indicative forms (like *vedo non vedo*, cf. (4b)). As regards the two instances of V_2SG.PRES.IND (*vedi non vedi*, lit. see.2SG.PRES NEG see.2SG.PRES), they are probably (meant to be) variants of *vedo non vedo*, or simply typos. See the figures in Tab. 1.

---

[6] We are indebted to Claudio Iacobini, who kindly gave us his personal annotated version of NVdB, that made the counting easy and fast.

Tab. 1: Category of the reduplicated form

Category	Subcategory	Tokens		Types	
N			4093 (89%)		685 (99%)
V	V_1SG.PRES.IND	514	516 (11%)	6	7 (1%)
	V_2SG.PRES.IND	2		1	
Total			4609		692

Half of the tokens display categorial ambiguity, meaning that the copy is interpretable either as a noun, identical to the first noun (thus yielding a N-*non*-N reduplicative structure), or as an adjective/participle that modifies a head noun (thus yielding a modifying structure rather than a reduplicative one). The high percentage of ambiguous cases is due to the very high frequency of *tessuto non tessuto*, which, as already observed in Section 3.1, is an ambiguous case, like *etica non etica* (see also (7b) below):

REDUPLICATIVE STRUCTURE					MODIFYING STRUCTURE		
tessuto$_N$	*non*	*tessuto*$_N$			*tessuto*$_N$	*non*	*tessuto*$_{PART\_PAST}$
fabric	NEG	fabric	vs.		fabric	NEG	woven
etica$_N$	*non*	*etica*$_N$			*etica*$_N$	*non*	*etica*$_{ADJ}$
ethics	NEG	ethics			ethics	NEG	ethic(al)

Fig. 5: Reduplicative structure vs. modifying structure

However, if one turns to types, the impact of ambiguous cases lowers considerably to only 21% of the cases, as illustrated in Tab. 2.

Tab. 2: Detail of ambiguous cases

Ambiguity	Subcategory	Tokens		Types	
no			2225 (48%)		545 (79%)
yes	ADJ	256	2384 (52%)	110	147 (21%)
	PART_PAST	2111		29	
	PART_PRES	17		8	
Total			4609		692

When the copy is unambiguous, it means that the reduplicated item is unequivocally a noun. When the copy is ambiguous, instead, the word can be a noun but also an adjective/participle. As Tab. 2 shows, if we look at tokens, we have a predominance of past participles (7a) (again, because of *tessuto non tessuto*), whereas if we consider types, in most cases we have a N/ADJ ambiguity (7b). Present participles, instead, are less frequent (7c).

(7) a. *Tisbe: quali sono le tue scelte faticose, scomode e costose? Non si capiscono dal tuo post, anzi, sembra che la tua sia quasi una **scelta non scelta**, tant'è che dici che ci sei nata così...* [itTenTen16]

'Tisbe: what are your strenuous, uncomfortable and expensive choices? You can't tell from your post, in fact, it sounds like yours is almost a '**non-choice**' [lit. **choice NEG choice/chosen**], so much so that you say you were born with it...'

b. *Berlusconi ha una sua coerenza e trasparenza. Può non piacere o disgustare. Ha una sua **etica non etica**, però tutti la conosciamo. Io temo più quelli che non conosco che quelli che conosco bene.* [itTenTen16]

'Berlusconi has his own consistency and transparency. One may dislike him or feel disgust towards him. He has his own **unethical/'fake' ethics** [lit. **ethics NEG ethics/ethical**], but we all know it. I fear more those I do not know than those I know well.'

c. [...] *quando parlavo con i ragazzi ad "X Factor" dicevo sempre che la verità non è stonata. [...] L'emozione non stona... Quanti **cantanti non cantanti** abbiamo che ci fanno venire la pelle d'oca?"* [itTenTen16]

'[...] when I was talking to the guys at "X Factor" I always said that truth is not out of tune. [...] Emotion is not out of tune... How many **not-proper-singers** [lit. **singers NEG singers/singing.PL**] do we have that give us goosebumps?"'

4. Analysis and discussion

The quantitative data illustrated in Section 3 show that the N-*non*-N pattern is a productive construction in contemporary Italian that is able to create new expressions. The lexical distribution of our dataset points to a situation where we have very few expressions that are more or less established in the lexicon (especially *tessuto non tessuto*, which is basically the only one recorded in dictionaries, cf. Section 4.2) and a very large array of new or occasional expressions formed according to this template, including proper names (like *Battiato non Battiato*, the name of a 1996 tribute album for singer-songwriter Franco Battiato).

N-*non*-N is employed typically with 'short' nouns (bi/trisyllabic). Even though we did not annotate semantic properties of nouns, we can report that the pattern can host a variety of semantic classes of nouns: both concrete (e.g., *libro non libro*, lit. book NEG book) and abstract (e.g., *bellezza non bellezza*, lit. beauty NEG beauty), both count (e.g., *preti non preti*, lit. priests NEG priests) and mass (e.g., *latte non latte*, lit. milk NEG milk). Sometimes hyphens are used to link the constituents, and the whole expressions are sometimes "scare-quoted".

The item that is reduplicated or copied is a noun in 99% of the cases in terms of types (coherently with our query on corpora), but we also found examples with a verbal form, mostly belonging to the *vedo non vedo* type. Both cases are 'not strictly predictable' in terms of formal properties and semantics. In the following we will focus on the nominal cases, which are the core of our present investigation.[7]

4.1 Unpredictable formal properties

As regards form, in Italian nouns are not found in this pattern in 'normal' syntax: we do find nouns followed by negation and another noun, but either (i) in juxtapositions with a contrastive function (like in (8a–b)), or (ii) in asyndetic coordination with additive function, like in (8c).[8] In both cases, the two parts are typically mediated by a comma before *non* in written language (which would plausibly correspond to a pause in spoken language), and most importantly identify two different referents.

(8) a. *Fa l'attore, non il cantante*
         'He's an actor, not a singer'

      b. *Producono macchine, non biciclette*
         'They produce cars, not bikes'

      c. *La manifestazione è aperta a scacchisti, non scacchisti, grandi e piccini.*
         'The event is open to chess players, non-chess players, young and old alike.'

---

[7] The verbal ones are even less predictable, by the way, since two identical verbal forms give rise to a complex nominal, thus displaying exocentricity.

[8] During the manual revision of the data automatically extracted from corpora, we identified instances of what seem to be other (peculiar) constructions. One is exemplified by (i), which is not approximation but rather a concise way to express that the 'question' is about using or not using béchamel (another way to express the same concept would be *besciamella sì besciamella no*, lit. béchamel yes béchamel no). Another one is exemplified by (ii), where *allenamento non allenamento* is detached from the main clause and emphasizes that the content of the main clause itself ('tomorrow evening he must be in Bologna') cannot be changed, *allenamento* 'training' being irrelevant (this construction would however sound better with a disjunction: *allenamento o non allenamento*, lit. training or NEG training).
(i) *Sulla questione **besciamella non besciamella**, io ce la metto ma io poi ovviamente non faccio testo* [itTenTen16]
     'About the question '**should-we-use-béchamel-or-not**' [lit. **béchamel NEG béchamel**], I use it, but my opinion is not worth tuppence'
(ii) [...] ***allenamento non allenamento** lui domani sera anche sul tardi deve essere a Bologna* [...] [itTenTen16]
     '**training or not** [lit. **training NEG training**], tomorrow evening, even late (evening), he must be in Bologna'

Differently from these cases, our N-*non*-N pattern forms a whole unit (sometimes marked by the use of hyphens or scare quotes, to keep the multiword expression together) with a unique referent, there is no pause or comma within the expression, and, most importantly, the second noun *must* be identical to the first. The latter property keeps it apart especially from cases like (8a–b). Indeed, contrastive juxtaposition with identical nouns would result in a contradictory, inappropriate utterance (#*Producono macchine, non macchine* 'They produce cars, not cars'), unless some (prosodic or gestural) emphasis is added to one of the two nouns, thus making them different, not identical items (e.g., *Producono MACCHINE, non macchine*).

Another peculiar formal property is that the second noun in the N-*non*-N pattern must be bare (e.g., *sapone non sapone* vs. **sapone non il/un/del sapone*, lit. soap NEG the/a/some soap), directly attached to *non*, even when the first noun is not: this would not be the case in contrastive juxtapositions, where the definiteness of the second noun mirrors the definiteness of the first noun (cf. **Fa l'attore, non Ø cantante*, with no article *il* 'the' before *cantante* 'singer').

Finally, the N-*non*-N sequence cannot be interrupted, not only by determiners, as just illustrated, but by any other element (e.g., *sapone non sapone* vs. **sapone profumato non sapone*, lit. soap scented NEG soap).

4.2 Unpredictable semantic properties

As regards semantics, N-*non*-N expressions are also unpredictable because they convey approximation of a concept. What is approximated is the meaning of N, whose core properties are called into questions. Let us take the following examples.

(9)   a.   *Casa nuova, vita nuova e quind... [sic] inviti a cena! Ovviamente sono impazientissima di tornare a cucinare e fare un po' di cene serie, nel mentre però, quando ci vuole ci vuole, e così ho fatto una* **cena non cena**, *insomma un invito a pane e salame, che poi proprio proprio pane e salami [sic] non è stato.* [itTenTen16]

       'New house, new life and therefore.... dinner invitations! Obviously I'm eager to get back to cooking and make some serious dinners, but in the meantime, when you need it you need it, so I made a 'non-dinner' [lit. **dinner NEG dinner**], in short, a bread-and-salami invitation, which was not so much bread-and-salami after all.'

b. *Zarazà: tradizione e modernità si mescolano per un mix affascinante e imperdibile. Uno dei piatti tipici è la famosa "**carbonara non carbonara**". [itTenTen16]*

'Zarazà: tradition and modernity intertwine for a fascinating and unmissable mix. One of the typical dishes is the famous "**carbonara NEG carbonara**".'

c. *Il riso si cuocerà nel tempo indicato sulla confezione, generalmente richiede 40 minuti di cottura. Il metodo della **cottura non cottura** è utile per sfruttare il tempo impegnato altrove per cuocere il riso.* [itTenTen16]

'The rice will cook in the time indicated on the package, generally requiring 40 minutes of cooking time. The 'non-cooking' [lit. **cooking NEG cooking**] method is useful for using the time committed elsewhere to cook the rice.'

d. *La legge che prometteva la chiusura definitiva degli ospedali psichiatrici giudiziari, quella specie di '**carcere non carcere**' per i malati di mente che si fossero macchiati di un crimine, è entrata definitivamente in vigore lo scorso 31 marzo.* [CORIS]

'The law that promised the definitive closure of forensic psychiatric hospitals, that sort of 'non-jail' [lit. **jail NEG jail**] for the mentally ill who are guilty of a crime, finally came into force on March 31.'

e. *Noi ci mettiamo il parco, la voglia e la medaglia. Voi i polpacci. Torna la **corsa non corsa**, che fa contenti tutti: mamme, papà, bambini, musicisti e scansafatiche. 6 chilometri di festa che altro non sono che una scusa per passare una domenica nel parco e ricordarsi che se si suda tutti assieme ci si sente meglio.* [itTenTen16]

'We'll throw in the park, the desire and the medal. You put your calves. Back to the 'non-running race' [lit. **race NEG race**], which makes everyone happy: moms, dads, kids, musicians and slackers. 6 kilometers of celebration that are nothing but an excuse to spend a Sunday in the park and remember that if you sweat all together you feel better.'

In (9a), a *cena non cena* (lit. dinner NEG dinner) identifies a kind of dinner which is not a full-fledged dinner, not apt for inviting people at home, which is an occasion that requires a 'proper' dinner, not a bread-and-salami frugal meal. *Carbonara non carbonara* in (9b) is the name of a special dish included in the menu of the Zarazà restaurant in Frascati: it's a 'carbonara' (a typical Roman recipe for pasta with eggs and bacon) with no eggs, so in this case a key ingredient is missing. In (9c) we find another food-related example: *cottura non cottura* here is an 'alternative' cooking method for rice that consists in letting it boil for a few minutes, turning off the flame (without uncovering the pot) and then letting it quietly cook by itself. In (9d) *carcere non carcere* is the expression the writer chooses for defining forensic psychiatric hospitals, comparing their status to that of a jail while at the same time

recognizing they are not officially jails, as also hinted at by the approximator *specie di* 'sort of' that introduces the expression (a 'redundancy' that we find in a few other examples). Finally, the event described in (9e) (the 'Magnolia Run' in Milan) is not a proper race since the true, ultimate goal is doing physical activity and spending quality time together rather than competing and winning.

As is clear from these examples, the semantics emerging from these expressions can be quite complex and specific, and is not the result of the sum of its parts: a literal reading of the parts would actually produce a contradiction; this triggers an unliteral interpretation which is not always easy to predict and that heavily relies on context. This is also testified by the fact that one and the same expression (not established in the lexicon) may carry (slightly) different meanings in different contexts. Take *canzone non canzone* in the following two examples:

(10)   a.   *Alla 57° edizione del Festival di Sanremo presenta "Ti regalerò una rosa". Una* **canzone non canzone** *in forma di lettera lacerante e commovente, microstoria di quel microuniverso della follia che tanto lo appassiona.* [itTenTen16]

        'At the 57th edition of the Sanremo Festival he presented "Ti regalerò una rosa". An unconventional song [lit. **song NEG song**] in the form of a lacerating and moving letter, a micro-story of that micro-universe of madness that fascinates him so much.'

   b.   *Un susseguirsi di brani non pianificati,* **canzoni-non-canzoni** *dilatate e contorte, come indagini radicali sulle possibilità del suono e del linguaggio.* [itTenTen16]

        'A succession of unplanned pieces, dilated and twisted 'non-songs' [lit. **songs-NEG-songs**], like radical investigations into the possibilities of sound and language.'

In (10a) the approximation concerns the fact that "Ti regalerò una rosa" is not a canonical Sanremo (the traditional Italian Music Festival) song in terms of lyrics, since it resembles a 'lacerating and moving letter'. In (10b) the same term is used to identify songs which are unplanned, dilated and twisted. So, in both cases we have some sort of non-canonical song, but the reason why it's non-canonical is different, pointing to different classes of objects.

Variation in context may be displayed also by expressions that do have a more established meaning: *colore non colore*, for instance, occurs 73 times in our dataset and in the vast majority of cases it refers to either 'white', 'black' or 'grey'. However, we also find more creative uses like (11), where the writer is reviewing a pink-like (cosmetic) illuminating powder, by describing it as very neutral and delicate, almost uncolored.

(11)   *La cialda è molto rosa, ma questo highlighter è molto meno rosa di ciò che sembra, è un **colore non colore**, stupendo, e la sua luminosità è pari quasi al caro Mary Lou citato prima.* [itTenTen16]

'The wafer is very pink, but this highlighter is much less pink than it looks, it's a '**non-color**' [lit. color NEG color], gorgeous, and its brightness is almost equal to the dear Mary Lou mentioned earlier.'

So, the kind of semantic operation the N-*non*-N pattern performs on N is one of approximation. More specifically, we claim it conveys 'non-prototypicality' by building a concept which is a deviating or peripheral instance of the category N, for some reason to be reconstructed by the reader/interlocutor. This 'deviating reason' is not always transparent and easy to get, and it may change from context to context (like for other approximating markers). Possible deviating reasons are:[9]

(i)   the lack of one or more salient/defining properties of N (see *cereale non cereale*, lit. cereal NEG cereal, referring to food plants like *grano saraceno* 'buckwheat', which – despite the name – is not closely related to wheat and is not a cereal);[10]

(ii)   the lack of one or more salient ingredients/parts of N (see the *carbonara* without eggs example in (9b));

(iii)   the lack of a shared function/purpose with N (see the *corsa non corsa* example in (9e));

(iv)   the lack of the expected/usual processes or factors involved to bring N into existence, despite having a similar function (see *sapone non sapone*, example (2), Section 1).

These reasons are reminiscent of Pustejovsky's (1995: 76) "qualia structure", which "specifies four essential aspects of a word's meaning (or *qualia*)", inspired by Moravcsik's (1975) interpretation of Aristotle's notion of modes of explanations. Deviating reason (i) would refer to Pustejovsky's[11] Formal quale, "which distinguishes the object within a larger domain" (think of properties like shape, magnitude, color, etc.); (ii) hints at the Constitutive quale, which specifies the "relation between an object and its constituents" or parts; (iii) is reminiscent of the Telic quale, which coincides with the "[p]urpose or function of the object"; and finally (iv) calls for the Agentive quale, namely the "[f]actors involved

---

[9] The list that follows is the result of a cursory observation of our data and is not meant to be exhaustive. However, we believe it might pave the way to a more fine-grained typology of approximative values (cf. Masini, Norde & Van Goethem 2023).

[10] Interestingly, these plants are called *pseudocereals* in English.

[11] All qualia quotations in the rest of this paragraph are from Pustejovsky (1995: 85–86).

in the origin or 'bringing about' of an object" (e.g., distinguishing between natural kinds and artifacts).

The presence of one or more 'deviations' might end up challenging the status of the N-*non*-N expression as a 'legitimate' type/instance of N: is a *sapone non sapone* a kind of *sapone*? The answer is no if we consider a soap an object that is obtained through saponification only (strict reading). The answer may be yes if we intend soap, in a broader sense, as a cleaning detergent for the body (loose reading).

The item we obtain is generally created on-the-go, giving rise to an ad hoc expression, but some of the items in our dataset became somehow conventionalised, although there is basically no record of these expressions in dictionaries (see Section 4.4).

4.3 N-*non*-N as a construction

We believe that the 'not strictly predictable' formal and semantic properties of the N-*non*-N pattern observed in Sections 4.1 and 4.2, respectively, speak in favor of the constructional status of this pattern, whose characteristics may be captured by a semi-schematic construction (where only *non* is lexically specified) that can be sketched as follows:

| FORM | PHON: preference for $\sigma\sigma(\sigma)$ Ns + prosody (no pause) |
| | MORPH/SYN: $[[N1]_j \, non \, [N2]_j]_{Ni}$ where N1=N2 |

| MEANING | SEM: $[\text{NON-PROTOTYPICAL SEM}_j]_i$ |
| | PRAG/DISC: - |
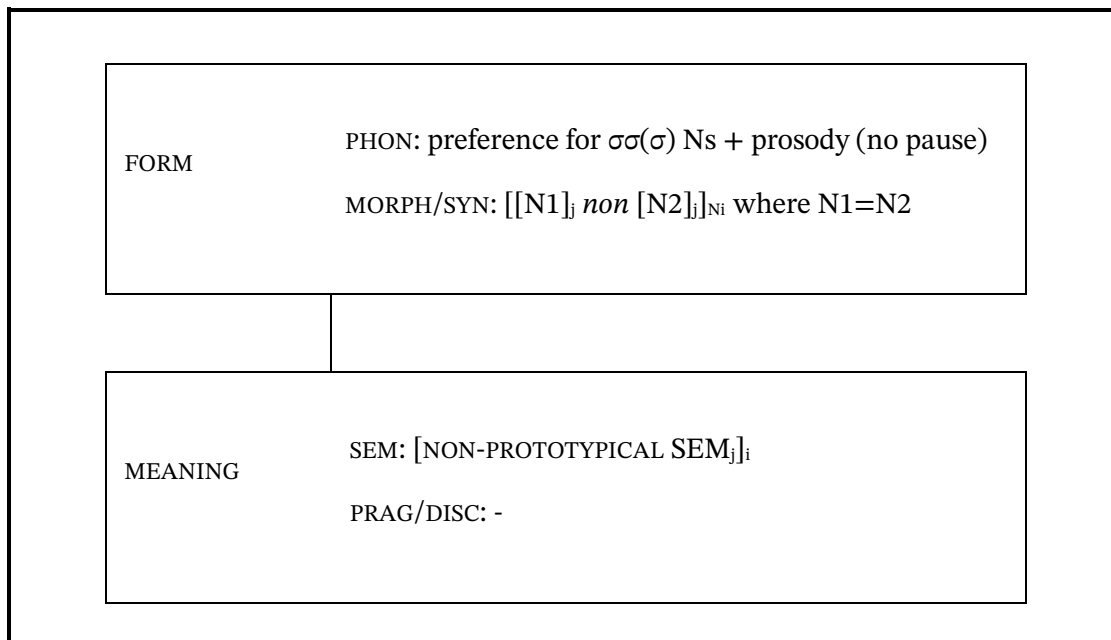
Fig. 6: The N-*non*-N construction

Note that the semantics of the construction, at this level, is quite abstract: the exact meaning conveyed by specific constructs (i.e., instantiations of this construction) can only be generated in context and can vary from time to time. In other words, the construction

in Fig. 6 does not state 'how' exactly non-prototypical the meaning of N will be (which aspect of meaning, or 'quale', will be 'deviant' with respect to the prototype), leaving a certain degree of flexibility that can be exploited creatively by the speakers.

As sketchily represented in Fig. 7, the maximally abstract construction in Fig. 6 can be instatiated both by lexically specified constructions (i.e. the few stored N-*non*-N expressions with conventionalised semantics, e.g., *tessuto non tessuto*) and by constructs, namely new extemporaneous coinings whose exact meaning is context-generated and may also vary from case to case by virtue of the semantic flexibility of the mother construction.
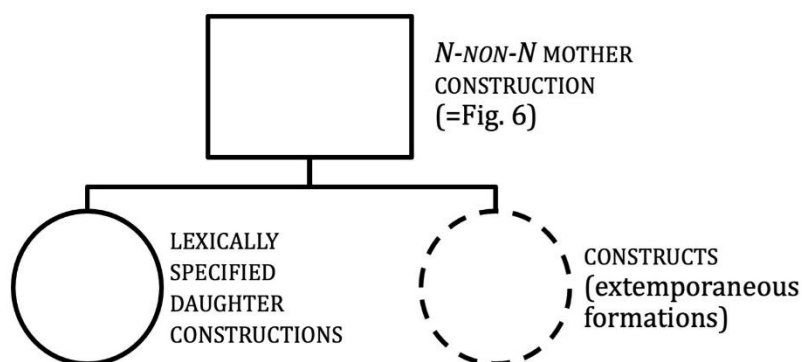
Fig. 7: The network of the N-*non*-N construction

Expressions such as *colore non colore*, which have a fairly stable meaning (however vague) but can also take on more extemporaneous meanings (cf. (11)), may find an explanation in the different possibilities of instantiation: whereas the semantically stable *colore non colore* may be a candidate for a lexically specified daughter construction, an extemporaneous *colore non colore* formation like (11) would be a construct, namely a direct, creative instatiation of the mother constructions.

We did not encode any genre/textual-related information into the construction because it is used in a variety of contexts. However, we may observe that it seems to be employed quite effectively to create terms for commercial products. Another (non unrelated) domain that seems to accommodate these formations quite easily is cooking and food. Both kinds of context probably rely on the expressiveness of these formations.

Finally, we decided not to encode any specific connotation. Sometimes, a pejorative connotation emerges (like for *cena non cena* in (9a) or *carcere non carcere* in (9d)), as quite expected in approximation. But in many instances there is no obvious derogatory effect: many expressions are quite neutral (like *carbonara non carbonara* in (9b)) and in some

cases they even convey, on the contrary, positive concepts (like *cottura non cottura* in (9c) or *corsa non corsa* in (9e)).

4.4 What about ambiguous cases?

As we said, a subset of nouns used in our construction (21% of types, including the very frequent *tessuto non tessuto*) are inherently ambiguous, since they can also be interpreted as adjectives/participles. This characteristic makes the analysis of these expressions uncertain: one may claim that these are not cases of reduplication because the second 'noun' is *not* a copy of the first noun but something different, namely an adjective/participle that modifies the first (head) noun. The two possible analyses may be sketched as follows (cf. also Fig. 5, Section 3.3):

(12) a. Reduplicative analysis: N-*non*-N$_{RED}$
      b. Modifying analysis: N-*non*-ADJ/PART

According to the modifying analysis, we are in front of a much less peculiar structure (cf. Section 4.1), namely, simply, a N followed by an adjective/participle. Also from a semantic point of view, the construction would gain in predictability and interpretability. If we read *tessuto non tessuto* as unwoven fabric, it becomes much easier to compute the overall meaning: it still conveys 'non prototypicality', in a way, but we do not need to guess the 'deviating reason' because it is explicitly stated by the participle (the object is not brought into existence by weaving); so, the 'non prototypicality' effect would be a by-product. In this respect, it is worth noting that ambiguous cases have a lower percentage of examples displaying hyphens and scare quotes (interpretable as orthographic cues to signal 'special' cases; see Section 3.2) with respect to unambiguous cases: ambiguous examples with hyphens and scare quotes amount to 9.1% and 7.1% (respectively) of total ambiguous cases, whereas unambiguous examples with hyphens and scare quotes amount to 28.6% and 17.1% (respectively) of total unambiguous cases.

It is difficult to find concrete evidence pro or against one of the two analyses in (12), because the items are genuinely ambiguous and both readings may be plausible in many cases. Let us take *etica non etica* (7b) (Section 3.3), repeated here as (13):

(13)　*Berlusconi ha una sua coerenza e trasparenza. Può non piacere o disgustare. Ha una sua **etica non etica**, però tutti la conosciamo. Io temo più quelli che non conosco che quelli che conosco bene.* [itTenTen16]

'Berlusconi has his own consistency and transparency. One may dislike him or feel disgust towards him. He has his own **unethical**/'**fake**' **ethics** [lit. **ethics NEG ethics**/**ethical**], but we all know it. I fear more those I do not know than those I know well.'

We may interpret *etica non etica* as 'unethical ethics' (modifying analysis), namely an ethics that is not morally correct (as it should be), or as a 'fake' or not genuine ethics (under the reduplicative analysis), possibly some sort of opportunistic or personalistic ethics. It is not trivial to pick one interpretation over the other, because they both hint at some (form of) non-prototypical ethics.

In this respect, the case of *tessuto non tessuto* deserves a broader discussion. On the one hand, as we already said, *tessuto non tessuto* is a non-prototypical *tessuto* 'fabric' irrespective of the analysis we choose. What happens if we read the second *tessuto* as a participle is that we get immediate access to *why* it is non-prototypical (and this would be one reason to regard this analysis as preferable, in fact), whereas in the other case it remains more vague and open to interpretation. How speakers actually perceive these expressions, however, is not a trivial question: even the two authors of this paper had opposite perceptions of *tessuto non tessuto* before starting the analysis. So, behavioral data is needed to better understand this point.

On the other hand, it is quite important to clarify the status of *tessuto non tessuto* since this specific item is most likely the 'leader expression' that paved the way to the emergence of the abstract N-*non*-N construction, given that it is by far the most frequent and established, and probably the oldest attested. *Tessuto non tessuto* is recorded as a main (multiword) entry in the *Nuovo De Mauro* online dictionary[12] and as a subentry of *tessuto* in the *il Sabatini Coletti* dictionary, but no first attestation date is provided. Both *il Sabatini Coletti* dictionary and *lo Zingarelli* dictionary have a main entry for *non tessuto* (carrying the very same meaning as *tessuto non tessuto*) with 1978 as a date of attestation.[13] In the *tessuto* entry of the Treccani Enciclopedia online,[14] we find that *non tessuto* is also called *tessuto non tessuto* (abbreviated as TNT). Finally, the full string *tessuto non tessuto* is first

---

[12] https://dizionario.internazionale.it/parola/tessuto-non-tessuto (accessed 6 March 2023).

[13] Nuovo De Mauro also has an entry for *non tessuto*, which however refers to the entry for *tessuto non tessuto*.

[14] https://www.treccani.it/enciclopedia/tessuto (accessed 6 March 2023).

attested in 1989 in the diachronic newspaper corpus *Archivio La Stampa*.[15] These pieces of information seem to suggest that there are two equivalent terms: *non tessuto* and *tessuto non tessuto*. The first has an earlier attestation, but the second one is much more used nowadays: a search on *itTenTen16* revealed that the occurrences of *non tessuto* amount to around 400, versus the 1966 occurrences of *tessuto non tessuto*. So, *non tessuto* appears first but is subsequently accompanied by (and superseded, in actual use, by) *tessuto non tessuto*, although most dictionaries do not reflect its primacy.

Note that also *non tessuto* is structurally ambiguous, since it can be perceived as a negated participle ('not woven/unwoven', which is probably the original, intended reading) but also as a noun preceded by prefixal *non* (cf. Iacobini 2004: 143; cf. also Dugas 2017 for the *non*-N construction in French), along the lines of other formations in Italian: the *lo Zingarelli* dictionary, for instance, contains (as subentries) *non luogo* (lit. NEG place), based on French *non-lieu* (a term introduced by French anthropologist Marc Augé in 1992), and *non notizia* (lit. NEG news, namely not a real news, because it's not new information or because it's irrelevant or lacks real content).[16] So, the second *tessuto* may have been perceived as a participle by some speakers (driven by its 'unwoven' semantics) but as a noun by other speakers (by analogy with other formations). This ambiguity is projected onto *tessuto non tessuto*, which emerges later but soon becomes a well-established and frequent term with its own acronym *TNT* (which often appears in fabric's labels). Our hypothesis is that *tessuto non tessuto* was (re)analysed as N-*non*-N by a sufficient number of speakers to eventually lead to the entrenchment of a more abstract N-*non*-N construction that could then be applied to *any* noun. The ambiguity, however, 'persists' with those nouns that *happen* to have an homophonous adjective or particle. In sum, what was initially a structurally ambiguous modifying structure (*tessuto non tessuto*) was reanalysed and then 'went viral'.

In support of this hypothesis, we observe two synchronic facts.

The first is that, if we look at the productivity of ambiguous vs. unambiguous cases, what we get is that the truly productive configuration is the unambiguous one, namely the one

---

[15] http://www.archiviolastampa.it/ (accessed 6 March 2023).

[16] Note that our dataset contains *luogo non luogo* (lit. place NEG place) and *notizia non notizia* (lit. news NEG news), with the same meanings. The two strategies – *non*-N and N-*non*-N – are obviously closely intertwined: their relationship and competition deserves a separate study. In addition, it might be worth looking into the N-*senza*-N pattern (*senza* meaning 'without').

that calls for the N-*non*-N~RED~ analysis. Table 3 shows that TTR and PP are much higher for the subset of our data including unambiguous cases only (even higher if we look at nominal cases only, thus excluding *vedo non vedo*) with respect to the whole dataset and the subset of ambiguous cases. So, synchronically speaking, the truly productive construction is the one that can be unequivocally interpreted as N-*non*-N~RED~ (namely, the one schematized in Figure 6, Section 4.3).

Tab. 3: Productivity scores for different subsets of our data

Productivity scores	Ambiguous cases only	Whole dataset	Unambiguous cases only	Unambiguous cases only, nouns only
TTR	0.06	0.14	0.24	0.31
PP	0.03	0.08	0.14	0.18

The second is that, among ambiguous cases, we find examples where the ambiguity is just virtual, since the actual interpretation that we obtain is not a consequence of the adjectival/participial reading of the second item. In other words, the second item does not contribute its meaning as an adjective/participle. This is what happens in example (7c) (Section 3.3), repeated here as (14) (short version):

(14)   *Quanti **cantanti non cantanti** abbiamo che ci fanno venire la pelle d'oca?* [itTenTen16]

        'How many **not-proper-singers** [lit. **singers NEG singers/singing.PL**] do we have that give us goosebumps?'

Here the ambiguity lies in the fact that *cantante* in Italian is both 'singer' and 'singing' (present participle). In (14), *cantanti non cantanti* identifies a special, non-prototypical category of singers, but their deviant property is not the fact that they are 'non-singing' (they do sing!). So, the only possible gloss would be "singers NEG singers".

    Similar examples are offered in (15): in (15a) *agenti* means both 'agents' and 'acting.PL' (present participle) but *agenti non agenti* identifies non-prototypical agents who are such not because they are 'non-acting' but because they are abusive; in (15b), *gelati* means both 'ice creams' and 'frozen.PL' (past participle) but *gelati non gelati* identifies a category of frozen products different from proper ice creams, not unfrozen products.

(15) a. *Nella filiera del mercato immobiliare, in Italia, gli **agenti non agenti**, ovverosia quelli abusivi, generano annualmente un'evasione fiscale che la Fiaip, Federazione Italiana Agenti Immobiliari Professionali, stima in ben 450 [...].* [itTenTen16]

'In the supply chain of the real estate market in Italy, **non-agents** [lit. **agents NEG agents/acting.PL**], i.e. abusive agents, annually generate a tax evasion that Fiaip, Italian Federation of Professional Real Estate Agents, estimates at 450 […]'

b. *Via libera insomma a grattachecche, ghiaccioli e sorbetti, tutti "**gelati non gelati**" ipocalorici.* [itTenTen16]

'Green light to *grattachecche*, ice lollies and sorbets, all low-calorie "**ice cream-like products**" [lit. **ice_creams NEG ice_creams/frozen.PL**].'

These data show that, even when the reduplicated item is ambiguous, technically speaking, the N-*non*-ADJ/PART interpretation is not necessarily activated, it may remain latent. In all these examples, it is in fact excluded, to the benefit of the competing reduplicative analysis.

These two observations seem to strengthen the idea of a truly reduplicative construction being quite entrenched and productive. What is the role of the ambiguous cases, then?

One may wonder whether ambiguous cases could be treated as instantiations of a separate daughter construction of the N-*non*-N construction with a less vague and more transparent semantics and a lower degree of productivity. However, it is very unclear if positing such a sub-construction is actually justified or necessary to account for the synchronic situation we depicted. The ambiguous cases may well be instantiations either of 'normal syntax', where a noun is modified by a homophonous adjective/participle for the sake of expressiveness,[17] or simply instantiations of the posited N-*non*-N construction.

In sum, although ambiguous *tessuto non tessuto* has most likely played a major role in the development of the N-*non*-N construction, synchronically speaking ambiguous cases may just be an epiphenomenon. If this analysis holds, then the emergence of the N-*non*-N construction would be an intriguing case of microdiachronic constructionalisation (given the short time span) due to the entrenchment of an expression whose original, intended structure does not coincide with the structure of the new construction being created.

---

[17] Note that examples like (14)–(15), despite being 'ambiguous', would not be instantions of this 'normal syntax', but of the reduplicative construction, for the semantic reasons given above.

5. Conclusions

In this paper, we provided the first thorough description of the N-*non*-N discontinuous reduplication in Italian, based on corpus data. The N-*non*-N construction proved to be a productive and powerful device to approximate nouns in contemporary Italian. More specifically, the construction has a general function of 'non-prototypicality' and can produce expressions with complex and specific meanings that are generated in context, by deviating from the 'prototypical' N concept along different directions. Despite its productivity, very few N-*non*-N expressions are conventionalised, making this device similar to other approximative strategies and to evaluative morphology in general.

Although this study is primarily synchronic in nature, we speculated about the origin of this construction, since the leader expression in our dataset (*tessuto non tessuto*) has an ambiguous structure that allows for both a reduplicative and a modifying interpretation. We hypothesised that the N-*non*-N construction microdiachronically emerged through the entrenchment of *tessuto non tessuto*, which was (re)analysed as a N-*non*-N structure, perhaps also helped by the existence of an independent *non*-N pattern yeilding nouns.

Many questions remain unanswered and deserve further investigation. One is prosody. Our analysis is based on written language, but the tendency to host 'short' nouns may reflect some prosodic requirement that can only be inquired through the analysis of spoken data. The semantics of these expressions, especially in relation to the 'deviations' from the prototype (but also to the ambiguity issue), also needs to be explored more fully, by resorting to speakers' judgements. Finally, a comparison with other constructions conveying approximation in Italian (cf. Masini & Micheli 2020) is in order, to better understand the role of this specific strategy within the larger picture.

References

Baayen, Harald & Rochelle Lieber. 1991. Productivity and English derivation: A corpus-based study. *Linguistics* 29. 801–843.

Baayen, Harald. 2009. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling & Merja Kytö (eds.), *Corpus Linguistics. An International Handbook*, 900–919. Berlin: Mouton de Gruyter.

Blackings, Mairi & Nigel Fabb. 2003. *A Grammar of Ma'di*. Berlin: Mouton de Gruyter.

Booij, Geert. 2010. *Construction Morphology*. Oxford: Oxford University Press.

Dugas, Edwige. 2017. Constraints on the French [*non*-N] construction. In Nikos Koutsoukos, Jenny Audring & Francesca Masini (eds.), *Morphological Variation: Synchrony and Diachrony. Online Proceedings of the Eleventh Mediterranean Morphology Meeting (MMM11) Nicosia, Cyprus, 22–25 June 2017,* 14–25.

Ghomeshi Jila, Ray Jackendoff, Nicole Rosen & Kevin Russell. 2004. Contrastive focus reduplication in English (The salad-salad paper). *Natural Language & Linguistic Theory* 22. 307–357.

Goldberg, Adele. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Goldberg, Adele. 2006. *Constructions at Work*. Oxford: Oxford University Press.

Iacobini, Claudio. 2004. Prefissazione. In Maria Grossmann & Franz Rainer (eds.), *La formazione delle parole in italiano*, 97–163. Tübingen: Niemeyer.

Inkelas, Sharon. 2014. Non-concatenative derivation: Reduplication. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford Handbook of Derivational Morphology*, 169–189. Oxford: Oxford University Press.

Jukes, Anthony. 2006. *Makassarese (basa Mangkasara'): A Description of an Austronesian Language of South Sulawesi*. PhD dissertation. University of Melbourne.

ListTyp = Masini, Francesca, Simone Mattiola & Stefano Dei Rossi. 2021. *The ListTyp database*. In Elisabetta Fersini, Marco Passarotti & Viviana Patti (eds.), *Proceedings of the Eighth Italian Conference on Computational Linguistics*, paper 17. CEUR Workshop Proceedings: http://ceur-ws.org/Vol-3033/ (accessed 6 March 2023).

Masini, Francesca & Maria Silvia Micheli. 2020. The morphological expression of approximation: the emerging *simil-* construction in Italian. *Word Structure* 13(3). 371–402.

Masini, Francesca, Muriel Norde & Kristel Van Goethem. 2023. Approximation in morphology: A state of the art. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1). 1–26.

Mattiola, Simone & Alessandra Barotto. 2023. Nominal reduplication in cross-linguistic perspective. From PLURALITY to CHANGE OF REFERENTS' SPECIFICITY. *Studies in Language* 47(1). 135–189.

Mauri, Caterina & Francesca Masini. 2022. Diversity, discourse, diachrony: A converging evidence methodology for grammar emergence. In Miriam Voghera (ed.), *From Speaking to Grammar*, 101–150. Bern: Peter Lang.

Moravcsik, Edith. 1978. Reduplicative constructions. In Joseph H. Greenberg (ed.), *Universals of Human Language. Volume 3: Word Structure*. 297–334. Stanford: Stanford University Press.

Moravcsik, Julius. 1975. Aitia as generative factor in Aristotle's philosophy. *Dialogue* 14(4). 622–638.

Pustejovsky, James. 1995. *The Generative Lexicon*. Cambridge: The MIT Press.

Stolz, Thomas. 2018. (Non-)Canonical reduplication. In Aina Urdze (ed.), *Non-Prototypical Reduplication*, 201–278. Berlin: Mouton de Gruyter.

Velupillai, Viveka. 2012. *An Introduction to Linguistic Typology*. Amsterdam: John Benjamins.

WALS = Dryer, Matthew S. & Martin Haspelmath (eds.). 2013. *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info (accessed 18 July 2021).

Francesca Masini
Alma Mater Studiorum – Università di Bologna
Dipartimento di Lingue, Letterature e Culture Moderne
Via Cartoleria 5
40124 Bologna, Italy
francesca.masini@unibo.it

Jacopo Di Donato
Alma Mater Studiorum – Università di Bologna Alumnus
Via IV Novembre 6
86100 Campobasso, Italy
jacopodidonato@gmail.com

M. Silvia Micheli

The emergence of approximative values in the Italian prefixes *semi-* and *para-*

Abstract: This paper examines the diachrony of two neoclassical prefixes in Italian, namely *semi-* (from Latin *sēmi-* 'half') and *para-* (from Ancient Greek *parà* 'close to'), based on data extracted from corpora. The diachronic analysis reveals that until the 20th century, both the prefixes *semi-* and *para-* were mostly used in scientific terminology, where they conveyed their original meanings: degree and spatial proximity respectively. A cursory synchronic analysis based on data extracted from two newspaper corpora shows that in the last decades, *semi-* and *para-* have been productively used to express evaluative values (i.e., attenuation and approximation). The present study enriches the literature on the sources of approximating morphological markers and the relationship between attenuation and approximation, considered here as two distinct values placed along a continuum.

Keywords: approximation, attenuation, diachrony, Italian, prefixes, neoclassical elements

1. Introduction

This paper contributes to the literature on the morphological expression of approximation by analyzing two prefixes identified by Masini & Micheli (2020) among those morphological elements that express approximation in Contemporary Italian, namely *semi-* and *para-*. Both prefixes are of classical origin and became part of the Italian lexicon through learned loans from Latin and Ancient Greek (e.g., It. *semidio* 'demigod' < Lat. *semideus*; It. *parabola* 'parable' < Lat. *parabola* < A. Gr. *parabolé*).[1] Particularly, *para-* derives from the Ancient Greek preposition *pará* 'close to' (see Bortone 2010: 291; Luraghi 2003: 131–45), while *semi-* originated from Latin *sēmi-* 'half' (cf. Ancient Greek *hēmi-* and Sanskrit *sāmi-*; cf. also Italian *emi-* 'half'), attested in Latin as both compound constituent (e.g., *semianimis*

---

[1] It should be noted that, notwithstanding their classical origin, *semi-* and *para-* are generally considered as prefixes (not as combining forms) within the literature devoted to Italian word-formation (cf. Iacobini 2004: 88, 101), since they express relational meanings.

'dying', lit. half + soul) and free form (i.e., the adjective/noun *semis* 'half'). The two prefixes are used in present-day Italian, both in scientific terminology and everyday language, as respectively exemplified in (1) and (2).

(1) a. *parapleurite*
         PARA + pleuritis
         'parapleuritis'

     b. *semiperimetro*
         SEMI + perimeter
         'half the perimeter of an object or shape'

(2) a. *diploma para-universitario*
         certificate PARA + academic
         'university-level certificate awarded by an institution other than a university'

     b. *seminterrato*
         SEMI + buried
         'basement'

The words in (1) exemplify the use of *para-* and *semi-* in Italian scientific terminologies, specifically in medicine and geometry lexicon. In *parapleurite*, referring to 'an inflammation affecting the tissues close to the parietal pleura', *para-* conveys the original value 'close to, beside'; likewise, *semiperimetro* describes the exact half of a perimeter, consistently with the value of *semi-* attested in Latin (namely, 'half'). Both prefixes also appear in everyday language, as exemplified by *diploma para-universitario*, in which *para-* denotes 'comparable to X, but not at the same level', and *seminterrato* 'basement', where *semi-* means 'partially X'. These values have been already identified in reference studies on Italian prefixation, such as Iacobini (2004: 131, 153); however, as will be explored in more detail throughout the article, *para-* and *semi-* are also used by speakers to express evaluative meanings, i.e., approximation and attenuation. Notably, the emergence of an evaluative meaning in originally locative prefixes (such as *para-*) is also attested in Modern Greek, as demonstrated by Efthymiou, Fragaki & Markos (2015a). However, in contrast with the Italian prefix, Modern Greek *para-* conveys an intensifying value ('excess'), such as in *para-cimáme*, meaning 'to oversleep' (cf. also *iper-* in Efthymiou, Fragaki & Markos 2015b).

The goal of this study is twofold. On the one hand, it investigates the emergence of evaluative meanings in two neoclassical prefixes through a diachronic analysis based on corpus

data. In doing so, it enriches the literature on the sources of approximating morphological markers. On the other hand, this study deepens the semantic analysis of the domain of approximation by analyzing the relationship between approximation and attenuation: notably, the two values are frequently considered a single semantic function within the literature on evaluative morphology, as testified by the terminology used to refer to them, e.g., "attenuation/approximation" in Grandi (2002, 2017) and Rainer (2015),[2] and "approximation/reduction/attenuation" in Grandi & Körtvélyessy (2015: 9, 11). The analysis of the semantic functions performed by the prefix *semi-* will contribute to the discussion on the boundaries between approximation and attenuation.

The article is organized as follows. Section 2 discusses the data extraction methodology and the parameters of analysis. The results of the analysis are presented in Section 3; the diachronic analysis of each prefix is supplemented with a cursory investigation of its use in Contemporary Italian. In Section 4 the results of the analysis are discussed and some perspectives for future research are suggested.

2. Methodology

This investigation is mostly based on data extracted from CODIT (i.e., Corpus Diacronico dell'ITaliano 'Diachronic corpus of Italian'; see Micheli 2022a), i.e., a thirty million word diachronic corpus of written Italian that covers a period from the earliest attestations of the language to 1947.[3] The corpus is structured into five subcorpora (see Tab. 1), corresponding to five historical periods comprised between crucial dates for the history of Italian language and literature (cf. the periodization proposed by D'Achille 1990 and adopted in MIDIA, Iacobini, De Rosa & Schirato 2017). Each subcorpus includes texts belonging to six genres: essays, literary prose, poetry, letters, scientific texts, and theatre; the only exception is the first subcorpus, which does not include scientific texts.

---

[2] Particularly, according to Rainer (2015: 1346), the "approximative/attenuative" function does not represent an autonomous value, but can be ascribed to "intensification", which comprises not only a high degree but all degrees of intensity.

[3] The corpus is hosted by the Czech National Corpus website; it can be searched through the KonText interface available at this link: https://www.korpus.cz/kontext/query?corpname=codit (accessed 3 February 2023).

Tab. 1: CODIT: Structure and size

	1 (13th c.–1375)	2 (1376–1532)	3 (1533–1691)	4 (1692–1840)	5 (1841–1947)
Essays	1,545,178	575,328	1,914,096	1,057,300	1,167,999
Letters	28,023	1,120,431	982,128	1,582,788	1,312,679
Poetry	945,047	1,674,318	1,257,071	856,507	1,350,275
Literary prose	1,393,433	1,585,877	1,472,700	1,501,061	1,808,785
Scientific texts	0	536,028	641,160	719,599	647,945
Theatre	65,537	410,363	465,163	449,978	428,717
	3,977,218	5,902,345	6,732,318	6,167,233	6,716,400

Since the corpus contains raw texts, it has been queried through the Kontext interface for forms that start with *<para>* and *<semi>*. A manual check was conducted to exclude false positives (e.g., *paradiso* 'paradise', and *semina* 'seeding'); finally, the selected forms were manually lemmatized. Tab. 2 illustrates the number of types and tokens of the *semi-* and *para-* words collected for each subcorpus.

Tab. 2: Dataset of words containing *semi-* and *para-* extracted from CODIT

	1	2	3	4	5
Semi-	16 tokens; 3 types	39 tokens; 3 types	281 tokens; 6 types	56 tokens; 23 types	84 tokens; 30 types
Para-	29 tokens; 3 types	75 tokens; 4 types	264 tokens; 6 types	446 tokens; 7 types	323 tokens; 9 types

The last part of the investigation is devoted to a cursory analysis of the prefixes *para-* and *semi-* in present-day Italian, based on data extracted from the *la Repubblica* corpus (see Baroni et al. 2004), comprising newspaper articles published from 1985 to 2000, and from a more recent journalistic corpus, namely the Timestamped JSI web corpus (2014–2016), searched through the Sketch Engine interface. The analysis was based on a sample of 100 occurrences of the two prefixes extracted from each corpus; given the large size of the two corpora, the analysis was limited to cases where the prefix is linked to the base through the hyphen to automatically extract words wherein the strings *para-* and *semi-* represent a true prefix and not a false positive. Table 3 illustrates the number of types and tokens of the *semi-* and *para-* words extracted from the two corpora.

Tab. 3: Dataset of words containing *semi-* and *para-* extracted from the two corpora of Contemporary Italian

	la Repubblica corpus (1985–2000)	**Timestamped JSI web corpus (2014–2016)**
Semi-	100 tokens; 72 types	100 tokens; 70 types
Para-	100 tokens; 63 types	100 tokens; 69 types

The dataset obtained has been analyzed according to the following parameters. From a formal point of view, the complexity of the base and the lexical category of the base and of the output have been considered. Semantically, the values conveyed by *para-* and *semi-* have been investigated through a qualitative analysis of the context of each. In the diachronic analysis based on CODIT, the textual genre in which the words are attested has been also taken into account.

3. Analysis

3.1 Analysis of the prefix *semi-*

Although not having particularly high-frequency values, s*emi-* is attested in all subcorpora of CODIT.[4] The following table illustrates the distribution of types, tokens, and hapaxes among the five subcorpora.

Tab. 4: Type/token frequencies and the number of hapaxes of words including *semi-* within CODIT

semi-	**1 (13th c.–1375)**	**2 (1376–1532)**	**3 (1533–1691)**	**4 (1692–1840)**	**5 (1841–1947)**
Types	3	3	6	23	30
Tokens	16	39	281	56	84
Hapaxes	1	1	0	16	21

As shown in Table 4, the type frequency of words containing *semi-* increases over time. More specifically, in the first two subcorpora, *semi-* only occurs in learned loanwords from Latin, where it means 'half' (i.e., *semicircolo* 'half circle' from Lat. *semicirculus*) or 'partially/partial' (i.e., *semidio* 'demigod' and *semivivo* 'unconscious' from Lat. *semivivus*). The latter value is exemplified in (3).

---

[4] It should be noted that CODIT also records two words containing *emi-*, namely *emicrania* 'migraine' and *emisfero* 'hemisphere'; they are learned loanwords from Ancient Greek (through Late Latin) whose structure reflects that shown by compounds in Ancient Greek made up of two bound forms.

(3)	*E*	*co'*	*pugni*	*stretti,*	*dette*	*queste*	*parole,*
	And	with	fists	clenched	say.PTCP.PST	these	words

	cadde	***semiviva***	*nelle*	*loro*	*braccia* […]
	fall.3SG.PST	SEMI + alive	in the	their	arms

'With clenched fists, after saying these words, she fell unconscious in their arms'. (Boccaccio, *Filocolo*, 1336–38)

At this stage, the prefix *semi-* cannot yet be considered a productive prefix since the only attested words containing it are all loanwords from classical languages. This scenario partly changes in the third subcorpus (1533–1691), where *semi-* occurs in three (very frequent) terms belonging to the field of geometry, which are attested only within the subcorpus of scientific texts; these terms include *semidiametro* 'semi-diameter' (200 occ.), *semicerchio* 'half-circle' (29 occ.), *semiparabola* 'id.' (21 occ.). The high value of token frequency for this period is due to the high token frequency of *semidiametro* within two specific texts by Galileo Galilei, namely *Due lezioni intorno alla figura, sito e grandezza dell'Inferno di Dante* (1588) and *Dialogo sopra i due massimi sistemi del mondo* (1632). The use of the prefix *semi-* in combination with terms of geometry is already attested in Neo-Latin, as shown by a considerable number of lemmas recorded by Ramminger's *Neulateinische Wortliste* (e.g., *semicircumferetia* 'semicircle')[5] and the *Dictionary of Renaissance Latin* by René Hoven (e.g., *semiglobus* 'hemisphere'). Interestingly, both dictionaries also record non-specialized terms used in literary prose, in which *semi-* expresses the attenuative value that will also emerge in Italian a few centuries later (e.g., *semierror* 'half mistake' or 'slight mistake' and *semifractus* 'half broken' in Hoven's dictionary and *semicerto* 'half certain' or 'not really certain' and *semilutheranus* 'semi-Lutheran' in Ramminger's *Neulateinische Wortliste*).

In the last two subcorpora, the productivity of *semi-* increases, as shown by the higher number of types and the presence of hapaxes. Focusing on the period from 1692 to 1840, *semi-* appears in a limited number of words belonging to scientific terminologies (i.e., *semitono* 'half tone' in music and *semi-membranoso* 'a muscle consisting of one-third of membranous tendon' and *semitendinoso* 'a muscle that has a long tendon of insertion' in

---

[5] The Ramminger's *Neulateinische Wortliste* covers a period ranging from 1300 to 1700. It is freely available at the following link: http://nlw.renaessancestudier.org/words/start.htm (accessed 30 January 2023).

medicine),[6] where it can denote 'half' or 'partially'. Notably, it also appears in a significant number of hapaxes (16) attested in historical/philosophical essays, where it combines with non-specialized nouns and adjectives, as exemplified in (4).

(4)   a. | *E* | *di* | *questo* | ***semi-amore***, | *o* | *sia* |
          |-----|------|---------|------------------|-----|------|
          | And | of | this | SEMI + love | or | be.3SG.SUBJ |

minore	*odio*	*del*	*tiranno*	*pel*	*popolo* […]
reduced	hatred	of the	tyrant	for the	people

      'And about this partial love, namely reduced hatred, felt by the tyrant towards the people'. (Vittorio Alfieri, *Della tirannide*, 1777)

      b. | *Quasi* | *che* | *un* | *uomo* | *potesse* | *essere* | ***semi-innocente*** |
          |--------|-------|------|--------|-----------|----------|----------------------|
          | almost | that | a | man | can.3SG.SBJV | be.INF | SEMI + innocento |

o	***semi-reo***,	*cioè*	***semi-punibile***	*e*	***semi-assolvibile***.
or	SEMI + guilty	that is	SEMI + punishable	and	SEMI + acquittable

      'Almost as if a man could be partially innocent or partially guilty, that is partially punishable and partially acquittable'. (Cesare Beccaria, *Dei delitti e delle pene*, 1764)

In (4a), the meaning of *semi-amore* is explicitly defined by the author as a low degree of hatred felt by the tyrant toward the people; in this case, *semi-* refers to a reduced degree of love (with a euphemistic nuance). In (4b), the same value is conveyed in four words including *semi-*: the author discusses the possibility of a man being partially guilty and therefore both partially punishable and acquittable (which he finds absurd).

Similarly, in the last subcorpus, *semi-* mostly occurs in historical or political essays where it combines with a still restricted number of adjectives and nouns (i.e., 10 A, 7 N, and 8 N|A), where it refers to a reduced degree of a category (e.g., *semi-indipendenza* 'partial independence'; [*società*] *semi-organizzata* 'partially organized society'), as shown in (5).

(5) | *Il* | *popolo* | *che* | *insorge,* | *per* | *avere* | *forti* | *probabilità* |
      |------|----------|-------|-----------|-------|---------|---------|---------------|
      | the | people | that | rebel.3SG.PRS | for | have.INF | strong | possibility |

di vittoria,	*deve*	*già*	*godere*	*di*	*una*	***semi-***
of victory	must	already	enjoy.INF	of	a	SEMI-

---

[6] Notably, both *semitonus* and *semimembranosus* are already attested in Neo-Latin (see respectively Hoven's dictionary and Ramminger's *Neulateinische Wortliste*).

indipendenza,		*in maniera*	*che*	*una*	*parte*	*di*	*esso*
independence		in way	that	a	part	of	it

sia	*militarmente*	*bene*	*organizzata.*
be.3SG.SBJV	militarily	well	organized

'In order to have a strong chance of victory, the rising people must already enjoy a partial independence, so that a part of them is militarily well organized.' (Gaetano Mosca, *Elementi di scienza politica*, 1896–1923)

Notably, in two cases ([*libro*] *semi-aperto* 'partially open' and *semi-accettazione* 'partial acceptance'), *semi-* occurs in text genres that are closer to everyday language, namely literary prose and personal letters, as illustrated in (6) with the case of *semi-accettazione*.

(6)	*Un*	*intimo*	*senso*	*di*	**semi-accettazione,**	*come*	*se*	*il*
	a	intimate	sense	of	SEMI + acceptance	as	if	the

male	*che*	*impreca*	*fosse*	*in*	*natura*	*ed*
evil	that	curse.3SG.PRS	be.3SG.SUBJ	in	nature	and

inevitabile.
unavoidable

'[…] an intimate sense of partial acceptance, as if the evil he curses were natural and inevitable'. (Massimo D'Azeglio, *Racconti, leggende e ricordi della vita italiana*, 1856)

The example in (6) highlights that the prefix *semi-* is also used in everyday language to express a partial accomplishment of an event conveyed by the base (see Grandi & Iacobini 2008). However, such cases are still very marginal, at least in the written language. The emergence of evaluative values related to approximation can be therefore considered more recent and linked to the use of *semi-* in a broader range of texts/contexts and in combination with a wider variety of bases. A cursory analysis of data extracted from the *la Repubblica* corpus (1985–2000) and Timestamped JSI web corpus (2014–2016) allows to observe in which contexts *semi-* is used and what kind of evaluative values it conveys in everyday language. Starting with the formal properties, Tab. 5 illustrates how the occurrences of *semi-* are distributed with respect to the input and output lexical categories as well as the complexity of the base.

Tab. 5: Formal features of words containing *semi-* in Contemporary Italian

	la Repubblica corpus (1985–2000)	Timestamped JSI web corpus (2014–2016)
Input category		
Adjective	63 tokens; 46 types	81 tokens; 54 types
Noun	37 tokens; 27 types	19 tokens; 18 types
Output category		
Adjective	63 tokens; 46 types	81 tokens; 54 types
Noun	37 tokens; 27 types	19 tokens; 18 types
Complexity of the base		
Complex	63 tokens; 45 types	76 tokens; 51 types
Simple	37 tokens; 28 types	24 tokens; 20 types

The data extracted from the *la Repubblica* corpus and those extracted from the more recent Timestamped JSI corpus are comparable and show the same trends. Regarding the input/output lexical categories, in most examples, *semi-* is attached to adjectives (especially relational adjectives, e.g., [*ordine religioso*] *semi-monastico* 'SEMI + monastic religious order') to form adjectives and, secondarily, to nouns to form nouns (e.g., *semi-libertà* 'SEMI + freedom'). Contrary to what Masini & Micheli (2020: 382–383) observed with respect to *simil-*, the addition of *semi-* does not induce a change in the category of the base. As for the complexity of the bases, most bases are morphologically complex words (e.g., *semi-clandestinità* 'SEMI + secrecy', where *clandestinità* is a noun derived through the suffix *-ità*).

Turning to the semantics of words containing *semi-*, the analysis of each token from both corpora reveals that *semi-* can express the following values:

1. 'attenuation, a reduced degree of X-hood' (i.e., the value attested in the last subcorpus of CODIT, e.g., in *semi-accettazione* 'partial acceptance'). The speaker wants to emphasize the attenuated degree of the entity/quality/event expressed by the base (sometimes with a derogatory or euphemistic nuance): the *semi-* word still falls within the category expressed by the base but has attenuated properties (e.g., *semi-addormentato* 'partially asleep', 'SEMI + asleep'). In most cases, such as *semi-addormentato*, the attenuation is due to the expressive needs of the speaker and is strictly context-dependent; however, a certain number of words containing *semi-* that

indicate established categories[7] are attested (e.g., [*pistola*] *semi-automatica* 'semi-automatic pistol'). The function expressed by *semi-* in these cases can be defined as *kincategorization*, adopting the term proposed by Masini and Micheli (2020: 384) to describe similar uses of *simil-* (i.e., "Y results in a separate (established) entity closely related to X, in that it is defined starting from X and deviating from it").

2. 'resemblance/imitation' (i.e., values belonging to the domain of approximation according to Masini & Micheli 2020):[8] the speaker questions the inclusion of the word containing *semi-* within the category expressed by the base due to the lack of certain fundamental properties (e.g., *semi-relazione* 'pseudo-relationship', 'SEMI + relationship')

As common in the study of qualitative evaluative values (Grandi 2017), the interpretation of the evaluative constructions is not always easy since it requires understanding the speaker's purposes from the analysis of the context: some cases have therefore been marked as 'doubtful'. Tab. 6 sums up the quantitative results of the semantic annotation of each token in the two corpora.

Tab. 6: Semantic classification of words containing *semi-*

	la Repubblica **corpus (1985–2000)**	**Timestamped JSI web corpus (2014–2016)**
'attenuation'	58 tokens; 42 types	47 tokens; 41 types
of which 'kincategorization'	15 tokens; 9 types	24 tokens; 14 types
'resemblance/imitation'	24 tokens; 22 types	25 tokens; 21 types
doubtful cases	3 tokens; 3 types	4 tokens; 4 types

As already observed for formal features, the two corpora show a similar distribution of occurrences, suggesting that the functions performed by *semi-* are stable over the time period considered. The more common value conveyed by *semi-* in both corpora is attenuation; as demonstrated by the example in (7a), *semi-* expresses a reduced degree of the category conveyed by the base, which is in this case a darkness that the speaker describes as partial.

---

[7] By "words referring to established categories", we mean words referring to stable entities with specific characteristics that are usually part of a taxonomy (e.g., in [*pistola*] *semi-automatica*, *semi-automatica* indicates a specific type of gun with precise characteristics). On the other hand, the form *semi-addormentato* is created by the speaker in context for expressive needs and reflects his subjective perception.

[8] Specifically, 'resemblance' is defined as "Y refers to an entity that is merely similar to X without being an X" while 'imitation' is defined as "Y refers to an entity that is meant to imitate or reproduce X (without being a genuine X" (Masini & Micheli 2020: 383).

Whereas in the case of *semi-oscurità* the attenuation is linked to the specific context, *semi-presidenziale* in (7b) indicates a stable category, namely a specific type of government system.

(7) a.
La	***semi-oscurità***	*permette*		*di*	*nascondere*
The	SEMI+darkness	permit.3SG.PRS		of	hide.INF

i	*numerosi*	*occhi*	*lucidi.*
the	numerous	eyes	shining.

'The partial darkness permits hiding the many shining eyes'. (*la Repubblica* corpus)

b.
Noi	*per*	*esempio*	*siamo*	*per*	*un*	*sistema*
we	for	example	be.3SG.PRS	for	a	system

semi-presidenziale	*alla*	*francese.*
SEMI + presidential	to the	French.

'For example, we stand for a French-style semi-presidential system'.
(*la Repubblica* corpus)

Besides the attenuative function, the prefix *semi-* also expresses values ascribable to the domain of approximation (i.e., 'resemblance' and 'imitation') when the speaker questions the inclusion of an entity or a quality to the category expressed by the base, as represented in (8).

(8) a.
Tinder è	*sia*	*una*	*app*	*sia*	*il*	*relativo*
Tinder be.3SG.PRS	both	a	app	both	the	related

semi-social network, *ma*	*è*		*soprattutto*	*un*	*servizio*
SEMI + social media	but	be.3SG.PRS	above all	a	service

di	*dating*	*online.*
of	dating	online

'Tinder is both an app and the related near-social media, but it is above all an online dating system'. (Timestamped JSI web corpus)

b.
Parla ,	*e*	*sotto*	*una*	*cantilena*	***semi-lombarda***
speak.3SG.PRS	and	under	a	singsong	SEMI + Lombard

riecheggia	*un*	*inconfondibile*	*accento*	*meridionale.*
echo.3SG.PRS	an	unmistakable	accent	southern

'(She) speaks, and under a Lombard-like singsong echoes an unmistakable southern accent'. (*la Repubblica* corpus)

In both examples, *semi-* expresses the idea that the entity (or the quality) referred to by the speaker lacks certain prototypical properties of the category expressed by the base and is therefore outside it. In (8a), Tinder is defined by the speaker as a 'near-social media' because, despite showing resemblance in some respects, it differs from prototypical social media (e.g., Facebook), since it has been primarily designed to facilitate online dating. In (8b), the speaker describes a woman who speaks with an accent that has some traits of the Lombard variety but not to the point of sounding authentic; in fact, the speaker himself still recognizes the typical features of the southern variety. Moreover, it is not clear from the context whether the woman is pretending to speak with a Lombard accent or not; in fact, *semi-* allows both an interpretation in terms of 'imitation' (i.e., a Lombard-like accent) and an interpretation in terms of 'fakeness' (i.e., a fake-Lombard accent; see the notion of *disproximation* proposed by Cappelle, Daugs & Hartmann 2023)[9].

Interestingly, *semi-* also appears in combination with other morphological elements that express values pertaining to approximation. In the example given in (9), *semi-* combines with *quasi-* (lit. 'almost') to reinforce the speaker's judgement that the militants of the Five Star Movement pretend to have anti-EU positions.

(9) *Sappiamo   bene   che   i   pentastellati   non   sono*
     know.1PL.PRS well   that the 5SM supporters   not   be.3PL.PRS

     *molto   veritieri   nelle   loro   istanze   **semi-quasi-antiUe**.*
     very   truthful   in the   their   demands   SEMI + almost + antiEU

     'We know that the Five Star Movement militants are not very truthful in their semi-almost-antiEU demands'. (*la Repubblica* corpus)

To sum up, the cursory analysis carried out on the two corpora of Contemporary Italian revealed that the main function performed by the prefix *semi-* is attenuation, which speakers utilize to signal the reduced degree of a given entity or quality. In such cases, the use of *semi-* does not imply a speaker's judgement on whether the entity belongs to a particular category, but it is rather a measurement in terms of degree internal to the category conveyed by the base. Besides the attenuative function, *semi-* is used when speakers want to

---

[9] According to Masini & Micheli (2020: 383), 'fakeness' is the value expressed when "Y refers to an entity that is meant to imitate X (without being a genuine X) in order to be taken as an X".

emphasize that a certain entity or quality resembles or imitates the one expressed by the base but is not precisely ascribable to it. In those cases, *semi-* conveys values pertaining to the domain of approximation, i.e., 'resemblance', 'imitation', and 'fakeness'. The two functions (attenuation and approximation) are therefore distinguishable in that, in the former case, the speaker does not question the inclusion of an entity (or a quality) to a category, unlike in the latter case. Needless to say, the two functions are very close and can be seen as situated along a continuum extending from attenuation (when the new concept does not exhibit all the prototypical properties but is still part of the category) to approximation (when the new concept is close to but does not belong to the category expressed by the base, in that it lacks properties that are considered essential).

3.2 Analysis of the prefix *para-*

Compared to *semi-*, the prefix *para-* is less frequently attested in CODIT, as shown in Table 7.

Tab. 7: Type/token frequencies and the number of hapaxes of words containing *para-* within CODIT

para-	1 (13th c.–1375)	2 (1376–1532)	3 (1533–1691)	4 (1692–1840)	5 (1841–1947)
Types	3	4	6	7	9
Tokens	29	75	264	447	323
Hapaxes	1	2	1	1	2

The type frequency of *para-* is significantly low among all subcorpora: in particular, starting from Old Italian, *para-* occurs in learned loanwords from Latin that were probably semantically and formally opaque already at this stage, e.g., *parabola* 'parable' (from Lat. *parabŏla*) and *parafrasi* 'paraphrase' (from Lat. *paraphrăsis*). Over the following centuries, these words became very frequent in texts, leading to an increase in the token frequency of the prefix. In contrast to what was observed in the case of *semi-*, Neo-Latin dictionaries do not list many examples where *para-* is productively used to form specialized or non-specialized terms; the dictionaries include a number of Ancient Greek loanwords or words created following the Ancient Greek model (e.g., *paraphronesis* 'delirium' or 'madness', *paragrammatismus* 'replacing one letter with another', *paratypum* 'counterfeit money', *parachronismus* 'contradiction in the chronological order', etc.), where *para-* suggests an anomaly or a deviation with respect to the meaning conveyed by the base.

Only in the last two subcorpora of CODIT one can find words that do not have classical/Neo-Latin origins; these include *parametro* 'parameter' (from French *paramètre*), *paraselene* 'each of the luminous discs that appear on the left and the right of the moon' (lit. beside the moon), and *paratartarico* 'paratartaric (acid)'. The last two words belong to scientific terminologies and exemplify two different semantic values conveyed by *para-*. In *paraselene*, *para-* has the original locative value (i.e., 'proximity'), while in *paratartarico*, which is used to refer to a type of acid that has some (but not all) of the properties of tartaric acid, it expresses 'similarity'. In both cases, *para-* is used to categorize/identify a given entity based on another entity to which it is close or similar, thus performing a kincategorization function.

The data extracted from CODIT only suggest that *para-* is used in scientific terminologies without giving any information on its usage in non-scientific varieties, which can be considered absent or extremely limited in the period ranging from the 13th century to 1947.

Turning to the most recent period, the data extracted from the *la Repubblica* corpus and Timestamped JSI corpus allow us to observe how *para-* is used in present-day Italian. Starting with the analysis of the formal features (see Table 8), *para-* tends to be attached to adjectives to form adjectives. The adjectives used as bases are mostly relational (e.g., [*cellule*] *para-terroristiche* 'terroristic-like cell'), but instances of qualifying adjectives are still attested (e.g., [*coppie*] *para-famose* 'pseudo-famous couples').

Tab. 8: Formal features of words including *para-* in Contemporary Italian

	la Repubblica corpus (1985–2000)	Timestamped JSI web corpus (2014–2016)
Input category		
Adjective	86 tokens; 51 types	85 tokens; 58 types
Noun	12 tokens; 12 types	14 tokens; 11 types
Proper Noun	2 tokens; 2 types	1 token; 1 type
Output category		
Adjective	86 tokens; 51 types	86 tokens; 59 types
Noun	14 tokens; 14 types	14 tokens; 11 types
Complexity of the base		
Complex	93 tokens; 56 types	80 tokens; 53 types
Simple	7 tokens; 7 types	20 tokens; 17 types

In most cases, the use of *para-* does not lead to a change of lexical category; however, in the most recent corpus, when combined with a proper noun, *para-* results in an attributive construction, as depicted in (10).

(10)   *I*   *fotogiornalisti*   *che*   *scelgono*   *di*   *imprimere*
       the   photojournalist   REL   choose.3PL.PRS   of   impress.INF

       un'   *impronta*   **para-Instagram**   *ai*   *propri*   *reportage*
       a   mark   PARA-Instagram   to the   their own   reports

       'Photojournalists who choose to impress an Instagram-style mark on their reports'. (Timestamped JSI web corpus)

Although under very limited conditions, the prefix *para-* seems to have category-changing properties, similar to what was already observed in the case of *simil-* (see Masini & Micheli 2020: 383). As for the complexity of the bases, *para-* clearly tends to combine with morphologically complex bases (especially adjectives, e.g., *para-bancario* 'para-banking' where *bancario* is an adjective derived through the suffix *-ario*).

    With regard to semantics, the analysis of each token from both corpora reveals that *para-* can mostly convey the following values:

1. 'subordinate to X, unofficial X': this value appears in more established words in which the base is a noun referring to an institution, a social group, or a professional figure (e.g., [*gruppo*] *para-militare* 'paramilitary, referring to a group that is organized like an army but is not official and often not legal' and *paramedico* 'a person whose training is similar to that of a nurse and who helps to do medical work').

2. 'resembling X': in these cases, *para-* performs an approximating function in that it signals that the entity referred to by the speaker is close to a given category but not fully included in it (e.g., [*intimidazione*] *para-mafiosa* 'mafia-like intimidation'). In some contexts, *para-* may also convey a pejorative nuance (e.g., *para-processo* 'pseudo-trial').

Table 9 summarizes the quantitative results of the semantic annotation of each token in the two corpora.

Tab. 9: Semantic classification of words containing *para-*

	la Repubblica corpus (1985–2000)	Timestamped JSI web corpus (2014–2016)
'subordinate to X'	51 tokens; 20 types	44 tokens; 23 types
'resembling X'	49 tokens; 44 types	56 tokens; 48 types

The two functions are almost equally distributed among the occurrences analyzed. In particular, the value 'subordinate to X' appears in words that are more entrenched within the lexicon (e.g., *para-militare* 'paramilitary': 20 occurrences within the *la Repubblica* corpus and 10 occurrences in the Timestamped JSI web corpus). Turning to the approximating function, it is mostly attested in nonce words where it fulfills expressive needs related to the specific context; in such cases, the speaker wants to refer to something that resembles a certain category, as illustrated in (11) where *para-leopardiana* indicates a philosophical conclusion that is in the style of the Italian poet Giacomo Leopardi.

(11) *Squisitamente* **para-leopardiana** *la* *conclusione* *filosofica [...]*
  Purely PARA-Leopardian the conclusion philosophical

  '(It's) a typically Leopardian-like philosophical conclusion [...]' (*la Repubblica* corpus)

Although it is not easy to identify from the context, in some cases *para-* conveys a pejorative nuance, as shown in (12) where the speaker criticizes people who think they are health-conscious but actually are not.

(12) *Il resto* *sono* *balle, anche quelle dei* **para-salutisti**
  the rest be.3PL.PRS bullshit also those of the PARA-health-conscious

  che *ogni* *sei* *mesi* *scoprono* *che* *mangiare*
  rel every six months find_out.3PL.PRS that eat.inf

  dietetico *fa* *bene* *al* *cuore*
  dietetic make.3SG.PRS good to the heart

  'The rest is bullshit, even those of the pseudo-health-conscious who, every six months, find out that dietary eating is good for the heart'. (Timestamped JSI web corpus)

Finally, as already noted for *semi-*, *para-* may combine with other prefixes or prefixoids expressing values pertaining to approximation in order to reinforce the evaluative meaning and make the utterance vaguer (cf., 'intentional vagueness' in Voghera 2012), as shown in (13).

(13)   *Uno*   *spettacolo*   *di*   *cui*   *erano*   *fino*   *a*   *un'*   ora

       a   show   of   REL   be.3PL.PST   until   to   a   hour

       fa   *protagonisti*   *(o*   *quasi-,*   *pseudo-*   ***para-protagonisti)***

       ago   protagonists   or   almost   pseudo   PARA-protagonists

       'A show in which they were protagonists (or almost, pseudo-, para-protagonists) until an hour ago.' (*la Repubblica* corpus)

To sum up, this cursory analysis has shown that the prefix *para-* is productively used in everyday Italian; notably, it appears with a similar frequency both in words referring to established categories with the meaning 'subordinate to X' and in nonce words where it conveys the approximative value 'resembling X'. The relationship between the two values will be discussed in the following section.

4. Discussion and conclusions

This study has enriched the literature on the development of approximating values in prefixes of classical origins. By analyzing the morphological and semantic properties of words containing *semi-* and *para-* in both diachronic and synchronic corpora, we found that both prefixes originally pertain to specialized terminology where they perform the categorizing function of creating new terms in scientific taxonomies. Focusing on *semi-*, only from the 17th century onwards, it is also attested in historical/political essays and (more rarely) in literary prose and personal letters, where it expresses attenuation by referring to a reduced degree of a category. Both values (scientific categorization and attenuation) are already attested in Renaissance Neo-Latin, which, as other studies have already shown (see, e.g., Rainer 2008 on *neo-*), anticipates some of the semantic values with which prefixes of classical origin are used in Italian in the later centuries. The synchronic analysis based on two Contemporary Italian corpora has shown that the main function of *semi-* is attenuation, which is performed both in words attenuated by the speaker for subjective evaluations and in words denoting established categories of reduced degree compared to those expressed by the base. This function does not imply a real qualitative evaluation by the speaker, but it is rather a measurement in terms of degree, closer to what Grandi (2002) calls 'objective evaluation'. Interestingly, besides the attenuative function, *semi-* is also used in contexts where the speaker wants to emphasize that a certain entity/property/event exhibits some

features of a certain category but not enough to be part of it. In such cases, the prefix performs approximative functions such as 'resemblance' and 'imitation', in that it refers to entities that are close to a category but not part of it. Indeed, the analysis of the contexts of use of *semi-* suggests that attenuation and approximation can be regarded as two distinct domains lying along a continuum, ranging from the former, where the entity mentioned by the speaker is inside the category but is downgraded, to the latter, where the entity is close but outside the category.

Turning to the prefix *para-*, the data extracted from CODIT are very scarce and seem to indicate a limited use of the prefix during the earliest centuries of the history of Italian. In the oldest subcorpora, it is attested only in opaque learned loanwords from Ancient Greek, and even Neo-Latin dictionaries record mostly words of Ancient Greek origin or neoclassical compounds, where the prefix indicates an anomaly or a deviation with respect to the meaning conveyed by the base (e.g., *parachronismus* 'contradiction in the chronological order'). This value is presumably related to the original semantic functions performed by *para-* in Ancient Greek alongside the locative one (i.e., 'contrary to' and 'other than'). Within CODIT, the only attestations of *para-* being used productively are in the field of specialized terminologies, where it may convey the original locative meaning (e.g., *paraselene*, lit. 'beside the moon') or indicate similarity (e.g., *paratartarico* 'tartaric-like acid'). While *para-* is productively used only in specialized terminology within CODIT, the analysis of Contemporary Italian data indicates that it appears in everyday language both in frequent and well-established words, which refer to something or someone that is subordinate to another entity conveyed by the base (such as *para-militare* 'paramilitary'), and in nonce words where it serves an approximating function (i.e., the value 'resembling'). In both cases, *para-* indicates that an entity has some properties of a category exemplar but not all of them (or at least not the most prototypical ones), so it may be similar but of a lower level or just resemble it.

This investigation has shown that the original values of the two prefixes play a role in the development of specific semantic nuances related to evaluation and particularly approximation. In the case of the prefix *semi-*, its original value 'degree measurement' resulted in an attenuative function and, in some contexts, in approximative values related to resemblance and imitation, on the basis that what has attenuated properties resembles the

full entity or quality. As for *para-*, the original locative value has left its traces in the current meaning 'subordinate to X' since subordination presupposes a closeness between the two entities; moreover, what is subordinate to X resembles X but does not have all its qualities. Finally, this study confirms that, alongside similarity items (see Masini, Micheli & Huang 2018 for an overview), fake items (see Van Goethem, Norde & Masini 2021; Van Goethem & Norde 2020), and diminutives (see, e.g., Stosic & Amiot 2023), spatial proximity items and degree items are possible sources of approximating morphological markers.

To conclude, this paper has highlighted some relevant issues that are worthy of further investigation. First, this study focused on the semantics of the prefixes *para-* and *semi-*, while very little has been said about their (possible) pragmatic effects: further research should verify whether (and to what extent) the prefixes *para-* and *semi-* can be used by speakers to modify the illocutionary force of speech acts, following the morphopragmatic perspective outlined by Dressler & Merlini Barbaresi (1994) (see, e.g., Mattiello 2009 on the mitigating function performed by negative prefixes). Second, a formal issue needs to be deepened: this study confirms a preference for prefixation in the expression of approximation in Italian, as shown by the great number of approximating prefixes or prefixoids identified by Masini and Micheli (2020: 393–397; i.e., *finto- para-*, *semi-*, *simil-*, *quasi-*, *non-*, *mezzo-*) in contrast to a smaller number of suffixes conveying this function (i.e., *-oide*, *-esco*, *-eggiante*, see Micheli 2022b). This issue deserves further study with a deeper focus on morpho-phonological properties shown by the bases. Preliminarily, we merely note that suffixation shows more constraints in terms of combinability with morphological complex bases, while prefixes exhibit a greater autonomy from the bases and can therefore (presumably) represent the favored strategy for constructions created in the moment by speakers. Finally, the adoption of a cross-linguistic (particularly Europe-wide) perspective seems crucial in the study of neoclassical elements in order to verify whether the emergence of evaluative meanings is attested in other European languages.

References

Baroni, Marco, Silvia Bernardini, Federica Comastri, Lorenzo Piccioni, Alessandra Volpi, Guy Aston & Marco Mazzoleni. 2004. Introducing the "la Repubblica" corpus: A large, annotated, TEI(XML)-compliant corpus of newspaper Italian. *Proceedings of LREC 2004*. http://www.lrec-conf.org/proceedings/lrec2004/pdf/247.pdf (accessed 7 June 2022).

Bortone, Pietro. 2010. *Greek Prepositions: From Antiquity to the Present*. Oxford: Oxford University Press.

Cappelle, Bert, Robert Daugs & Stefan Hartmann. 2023. The English privative prefixes *near-*, *pseudo-*, *quasi-* and *sub-*: Approximation and 'disproximation'. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1). 52–75.

D'Achille, Paolo. 1990. *Sintassi del parlato e tradizione scritta della lingua italiana*. Roma: Bonacci.

Dressler, Wolfgang U. & Lavinia Merlini Barbaresi. 1994. *Morphopragmatics. Diminutives and Intensifiers in Italian, German, and Other Languages*. Berlin: De Gruyter.

Efthymiou, Angeliki, Georgia Fragaki & Angelos Markos. 2015a. Exploring the meaning and productivity of a polysemous prefix: The case of the Modern Greek prepositional prefix *para-*. *Acta Linguistica Hungarica* 62(4). 447–476.

Efthymiou, Angeliki, Georgia Fragaki & Angelos Markos. 2015b. Exploring the polysemy of the Modern Greek prefix *iper-*. *Morphology* 25(4). 411–438.

Grandi, Nicola. 2002. *Morfologie in contatto. Le costruzioni valutative nelle lingue del Mediterraneo*. Milano: FrancoAngeli.

Grandi, Nicola. 2017. Evaluatives in Morphology. In *Oxford Research Encyclopedia of Linguistics*. DOI: 10.1093/acrefore/9780199384655.013.250.

Grandi, Nicola & Claudio Iacobini. 2008. L'affissazione valutativa nei verbi dell'italiano. In Emanuela Cresti (ed.), *Nuove prospettive nello studio del lessico. Atti del IX Congresso Internazionale SILFI (Firenze, 15-17 giugno 2006)*, 483–490. Firenze: Firenze University Press.

Grandi, Nicola & Lívia Körtvélyessy. 2015. *The Edinburgh Handbook of Evaluative Morphology*. Edinburgh: Edinburgh University Press.

Iacobini, Claudio. 2004. Composizione con elementi neoclassici. In Maria Grossmann & Franz Rainer (eds.), *La formazione delle parole in italiano*, 69–95. Tübingen: Niemeyer.

Iacobini, Claudio, Aurelio De Rosa & Giovanna Schirato. 2017. Criteri e strategie di classificazione morfosintattica dei testi del corpus MIDIA. In Paolo D'Achille & Maria Grossmann (eds.), *Per la storia della formazione delle parole in italiano. Un nuovo corpus in rete (MIDIA) e nuove prospettive di studio*, 33–55. Firenze: Cesati Editore.

Luraghi, Silvia. 2003. *On the Meaning of Prepositions and Cases: The expression of Semantic Roles in Ancient Greek*. Amsterdam/Philadelphia: John Benjamins.

Masini, Francesca & M. Silvia Micheli. 2020. The morphological expression of approximation. The emerging *simil-* construction in Italian. *Word Structure* 13(3). 396–427. DOI: 10.3366/word.2020.0176.

Masini, Francesca, M. Silvia Micheli & Shanshan Huang. 2018. Attenuation and categorization: A cross-linguistic study of similarity items. Paper presented at the Workshop "Attenuated qualities in a cross-linguistic perspective", 51st Annual Meeting of the Societas Linguistica Europaea (SLE 51), Tallinn, 29 August–1 September.

Mattiello, Elisa. 2009. A morphopragmatic analysis of English and Italian negative prefixes. *Studi e Saggi Linguistici* XLVII. 125–156.

Micheli, M. Silvia (2022a). CODIT. A new resource for the study of Italian from a diachronic perspective. Design and applications in the morphological field. *Corpus* 23. *Special Issue "Corpus et Données en Morphologie"*. DOI: 10.4000/corpus.7306.

Micheli, M. Silvia (2022b). Italian approximative suffixes in diachrony. Poster presented at the 13th Mediterranean Morphology Meeting. Rhodes, 19–21 May.

Rainer, Franz. 2008. Neo- and neo-Latin. *Word Structure* 1(1). 53–64.
DOI: 10.3366/E1750124508000068.

Rainer, Franz. 2015. Intensification. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-formation. An International Handbook of the Languages of Europe*, vol. 2, 1339–1351. Berlin: De Gruyter.

Stosic, Dejan & Dany Amiot. 2023. Why does morphological augmentation not lead to approximation? The case of denominal evaluative nouns. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1). 177–203.

Van Goethem, Kristel & Muriel Norde. 2020. Extravagant "fake" morphemes in Dutch. Morphological productivity, semantic profiles and categorical flexibility. *Corpus Linguistics and Linguistic Theory* 16(3). 425–458.

Van Goethem, Kristel, Muriel Norde & Francesca Masini. 2021. The fate of 'pseudo-' words: A contrastive corpus-based analysis. Paper presented at UCCTS 2021 (Using Corpora in Contrastive and Translation Studies), Bertinoro, 9–11 September 2021.

M. Silvia Micheli
Università degli studi di Milano – Bicocca
Piazza dell'Ateneo Nuovo 1
20126 Milano, Italy
maria.micheli@unimib.it

Dejan Stosic & Dany Amiot

Can suffixal augmentation express approximation?
The case of denominal evaluative nouns

Abstract: The paper focuses on the relationships between so-called diminutive/augmentative morphology and approximation. As this topic is very broad, only the denominal nouns are studied in this article. Our study, based on data from French and Serbian, has a twofold aim: to answer the question asked in the title of this article and, the answer being negative, try to understand why only diminutive suffixes, and not augmentative ones, can express approximation. This should also allow us to understand the specificity of diminutive approximation. Firstly, we show that even though both types of evaluatives express non-conformity with respect to a norm, in the case of diminution the norm is not reached, whereas augmentation entails that the norm is not only reached but is exceeded. This suggests that approximation is fundamentally associated with a certain degree of deficiency, which is consistent with the literature on this topic. Secondly, we demonstrate that the main peculiarity of diminutive approximation is that it is measurative in nature as it is underlined by the orientational metaphor UP/DOWN and based on a set of gradable semantic dimensions such as SMALL/BIG, FEW/MUCH. This is not the case of other morphological devices that trigger approximative meanings (e.g., *pseudo-, -oïde*), whose role is to provide a global evaluation of the matching between a given referent and core members of the category expressed by the base.

Keywords: approximation, diminution, augmentation, evaluative morphology, categorization

1. Introduction

Diminutive and augmentative meanings are quite typical of evaluative morphology, whatever the language concerned (see, e.g., Jurafsky 1996; Grandi & Körtvélyessy 2015; Štekauer 2015; Martín Calvo 2022). Evaluative morphology fundamentally provides items that express an appreciation of existing conceptual categories (see for example Scalise 1986; Stump 1993; Bauer 1997; Fradin 2003; Dressler & Merlini Barberesi 1994; Stosic & Amiot 2019), regardless of the category involved: noun (*fille* 'girl'/*fillette* 'little girl'), verb (*voler* 'to fly'/*voleter* 'to fly'/'to flutter'), adjective (*propre* 'clean'/*propret* 'neat and tidy'), preposition

(Breton: *e-kichen* 'near'/*e-kichenig* 'very near'[1]), etc. In a previous article (Amiot & Stosic 2022) focusing on diminutive evaluative suffixation, we showed that, in French, approximation can only be expressed by a subset of these derivatives, which we called "strong approximative evaluatives". In accordance with other contributions from this volume, we do not limit approximation to one of the senses of diminution identified by Jurafsky (1996: 536, Tab. 2), alongside "small, young/offspring, female gender, small-type, imitation, intensity/exactness, individuation/partiality". This sense of approximation, called by other authors "attenuation", is considered by Jurafsky, and other authors after him, as characteristic of adjectival evaluation (*yellow*/*yellowish*). We rely on a broader definition of approximation. This is based on the definition of Mihatsch (2009: 100), who considers that the domain of approximation concerns "explicit linguistic means that serve to blur the categorical limits of the linguistic element they modify". These "explicit linguistic means" alluded to by Mihatsch are what Lakoff (1973) calls "hedges", i.e., "words whose meaning implicitly involve fuzziness-words whose job is to make things fuzzier or less fuzzy" (Lakoff 1973: 471; see also Prince et al. 1982; Kleiber 1987; Channell 1994; Gerhard-Krait & Vassiliadou 2017; Vassiliadou & Lammert eds. 2022). In evaluative morphology, whose scope encompasses a variety of lexical categories, the suffix can be this "explicit linguistic mean" (on this subject, see Amiot & Stosic 2022). Consequently, approximation is not taught as a simple attenuation of qualities, but as a range of strategies relating to fuzzy or vague representation of conceptual categories (for further discussion, see Introduction to this issue by Masini et al. 2023, and below).

In this article, we extend our research to suffixal augmentation by investigating whether it can also express approximative meanings. The answer to the question being negative, we will try to understand why, which should in turn give us additional information about approximation itself.

To do so, we proceed in several steps: we first try to give a definition of morphological approximation, but it will be rather restrictive because it fits our specific object: diminutive/augmentative suffixation (Section 2). Then, based on the results of our previous work (Amiot & Stosic 2022), we recall the constraints that enable these approximative meanings

---

[1] Example from Stump, 1993: 2. On the grammar of Breton, see Jouitteau (2009–2023), and on the suffix *-ig* in particular, see https://arbres.iker.cnrs.fr/index.php?title=-ig#base_adverbiale (accessed 13 December 2022).

to be constructed in French suffixation (Section 3). Since French has a very poor augmentative evaluative suffixation[2], Serbian, a Slavic language with a more fully developed (diminutive and augmentative) evaluative suffixation, will serve as a reference language. We chose to focus on nouns because, as shown by Grandi (2009: 47), from a typological point of view, nouns play a very central role in the formation of both diminutives and augmentatives and because this category seemed to be more suitable for raising the question of approximation, and thus the degree of membership to the conceptual category associated with the base.

Our analysis, based on a corpus of 814 nouns suffixed with one of the 13 suffixes that can express augmentation in Serbian (cf. Klajn 2003; Jovanović 2010; Bońkowski 2017), shows that evaluative augmentative derivatives, while involving the same formation principles as diminutive derivatives, never express approximation (Section 4). This observation leads us to ask the question of "why" (Section 5).

2. Morphological approximation: a working definition

The present study focuses on a set of evaluative suffixes from French and Serbian that form diminutive (French) and augmentative (Serbian) meanings. Here are some nominal examples:

French:

(1)   a.   *-et/-ette*: *île* 'island' → *îlet* 'small island, islet'; *savon* 'soap' → *savonnette* 'small bar of soap'

        b.   *-ot/-otte*: *bec* 'beak, kiss' → *bécot* 'small kiss/peck', *main* 'hand' → *menotte* 'tiny hand'

        c.   *-ille*: *brin* 'sprig, strand' → *brindille* 'twig, stalk'

---

[2] In French, the suffixes *-issime, -on* and *-asse*, more rarely *-âtre*, are sometimes considered as augmentative suffixes, however:
– while this article focuses on nominal suffixation, the suffix *-issime* is an adjectival suffix: *riche* 'rich' → *richissime* 'extremely rich'; *ridicule* 'ridiculous' → *ridiculissime* 'extremely ridiculous'. In French, nominal augmentation can be built by prefixation, especially with *hyper-* (N: *sensibilité* 'sensitivity' / *hypersensibilité* 'hypersensitivity'). Prefixal augmentation (and diminution) raises other complex questions that we could not address in this article.
– The suffixes *-asse* and *-on* do show an augmentative meaning in some nouns, e.g., *-asse*: *barque* 'rowing boat'/ *barcasse* 'big rowing boat', but also 'bad boat'; *-on*: *balle* 'ball' → *ballon* 'big ball', but these are usually borrowed nouns: *barcasse* is thus supposed to have been borrowed from Italian (*barcaccia*) or Provencal (*barcasso*), and *ballon*, from Italian (*pallone*). In general, nouns built with these suffixes have other interpretations: e.g., for *-on*, the diminution, with a notion of partitivity (*glace* 'ice' → *glaçon* 'ice cube'), and/or familiarity (*curé* 'vicar' → *cureton* 'vicar, informal'), or, for *-asse*, an essentially pejorative meaning (*vin* 'wine' → *vinasse* 'cheap wine, plonk').

Serbian:

(2)   a.   *-ina*: *kuća* 'house' → *kućetina* 'a very big (ugly) house'

      b.   *-čuga*: *bolest* 'illness' → *boleščuga* 'serious illness'

      c.   *-erda*: *ruka* 'hand' → *ručerda* 'a big odious hand'

Given the studies presented in this volume and our previous paper on morphological approximation (Amiot & Stosic 2022), we can observe that this device constructs at least two main types of approximative meanings, all of which, however, involve an evaluation with respect to a standard: meanings essentially related to fakeness, imitation, non-authenticity on the one hand, and meanings related to diminution on the other hand. The two concepts can overlap, but we will focus on the second concept, the only one represented by our suffixes.

An approximative derivative fundamentally expresses both resemblance and dissimilarity with respect to the norm; approximation thus involves categorizing – or at least an attempt to categorize – an entity in a pre-existing class represented by the base. This general mechanism of matching what is to be designated and the category that a word represents is also present in non-morphological strategies (cf. "hedges", Lakoff 1973; Prince et al. 1982; Mihatsch 2009):

(3)   Eng.   *I saw a very strange creature,* **_a kind of_** *bird...*

          Fra.   *Il a* **_une sorte d'engin_** *qui coupe tout...*
                 'He has a kind of device that cuts everything'

In all these cases there is an imperfect match, leading to a kind of fuzzy representation, which can be of two types generally speaking: referential or pragmatic.

3. Diminutive suffixation: The construction principles of approximative meaning

3.1 French approximative suffixation

In French, approximation can be expressed by (at least) two sorts of suffixes, as illustrated by examples (4)–(5) vs. (6):

– The first sort can be illustrated by the suffixe *-oïde* (4) and the affixoid *-forme* (5)[3]:

(4) a. *une roche granit**oïde*** 'a granitoid rock' is 'a rock whose texture is close to that of granite'

   b. *un état crétin**oïde*** 'a creninoid state' is 'a state reminiscent of cretinism, but less pronounced'

(5) a. *une substance géli**forme*** 'a gel-like substance' is 'a substance that looks like gel'

   b. *une boisson théi**forme*** 'a theiform drink', that is a 'drink reminiscent of tea, which is prepared in the same way as tea'

Adjectives in *-oïde* and *-forme* fundamentally express both resemblance and difference with respect to an entity that serves as a norm and that is represented by the base of the derivative.

– The second sort of suffixation is illustrated by the French diminutive suffixes, as in (6)[4]:

(6) *-et/-ette*: *réforme* 'reform' → *réformette* 'mini-reform with little value or scope'

   -âtre: *jaune* 'yellow' → *jaunâtre* 'yellowish'

   -ill(er): *sauter* 'to jump' → *sautiller* 'to hop', etc.

The very use of the term *réformette*, for example, means that the referent does not deserve the name of *réforme* (*c'est une réformette, pas une réforme!* 'it is a "réformette", not a reform!*): it looks like one but does not have the scope of a real reform. The use of the noun *reformette* seems to imply, from the speaker, a certain detachment based on a kind of metalinguistic judgment. As for *jaunâtre* 'yellowish', it means 'somewhat yellow', often with a pejorative value. The meanings built by the suffixations *-oïde* and *-forme* and those built by *-et, -âtre* or *-iller* are different, but in both types, a meaning of approximation is built from an implicit comparison between the prototype associated with the base: the derivative then expresses non-conformity with respect to this standard. In addition, *réformette* expresses a second type of non-conformity with a standard: it is pragmatically marked in that it does not belong to the standard vocabulary but to distinctive registers, especially informal ones with a pejorative value.

---

[3] On the analysis of *-form* as an affixoid, see Amiot & Dugas (2020).

[4] Besides *-et/-ette*, *-âtre* and *-ill(er)*, there are several other diminutive suffixes in French: *-ouill(er), -asse, -ot/ -ot(t)e, -eau*, etc. Authors who have studied evaluative suffixation in French agree that it is difficult to determine the exact number of suffixes (see for example Plénat 1999).

In our previous study, we focused on examples of type (6), which belong to evaluative morphology in the narrow sense. In this sort of morphology, the evaluation is accomplished through a set of binary oppositions that universally structure human conceptual representations, and that can be represented as in Fig. 1.
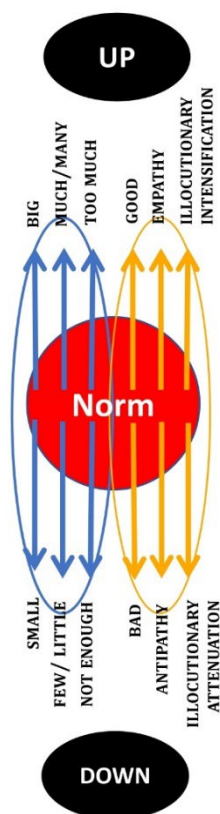
Fig. 1: Bipolarity of evaluative meanings[5]

Following Lakoff & Johnson (1980: 14–21), each axis can be defined as a system of values underlined by the orientational metaphor UP-DOWN that gives the conceptual domain at issue a specific spatial orientation. The authors argue that several fundamental conceptual domains are organized in terms of the UP-DOWN spatialization metaphor: GOOD IS UP/BAD IS DOWN, MORE IS UP/LESS IS DOWN, VIRTUE IS UP/DEPRAVITY IS DOWN, HAPPY IS UP/SAD IS DOWN. As for evaluative morphology, we showed in our previous work, on the basis of the relevant literature (Dressler & Merlini Barberesi 1994; Mel'čuk 1994; Grandi 2002; Fradin

---

[5] This representation is based on our previous studies (especially Amiot & Stosic 2019 and 2022, also Stosic & Amiot 2011) but has evolved so as to be consistent with the development of our analyses. As for the bipolarity, i.e., the distinction between the Referential Pole and the Pragmatic Pole, it is inspired by the work of Fradin (2003), taken up in Meunier et al. (2007). The same applies to the term "measurative", borrowed from Fradin and taken up, among others, by Efthymiou (2017).

2003; Fradin & Montermini 2009; Grandi & Körtvélyessy 2015 eds.; Štekauer 2015), that two sets of values can be distinguished, forming respectively two different poles: the Referential Pole (RP), in blue, which represents the set of interpretations involving a measurative dimension; it gives rise to measurative meanings of the type: SMALL/BIG, LITTLE/MUCH, NOT ENOUGH/TOO MUCH, i.e., the meanings of diminution, augmentation, excess, etc. The other axis, in yellow, represents the Pragmatic Pole (PP), which contains the set of interpretations involving the subjective or appreciative dimension, and gives rise to more qualitative meanings, closely related to speaker affect ((dis)satisfaction, (dis)approval…): GOOD/BAD, EMPATHY/ANTIPATHY, etc. In both cases, however, the evaluation is carried out from a standard inherent in the concept of the base, namely, the reference point (the "Norm" placed in the middle of the two axes) of the evaluation system. As for the arrows oriented in the two directions, toward the UP or toward the DOWN side, they represent the evaluation process itself that takes place in relation to N: excess or a large quantity, for example, are located on the upper part of the axis (UP). Conversely, smallness or insufficiency are located on the lower part of the axis (DOWN). This placement on either side of the zero point of the system on an oriented axis constitutes, in our opinion, the specificity of evaluative suffixation in the strict sense as it expresses diminution or augmentation. We shall see that this specificity has an influence on the capacity of suffixes to express approximation (Section 3.2), and on the type of approximation expressed by these suffixes (Section 5.3). Consequently, the UP-DOWN orientational metaphor can be considered as providing the basis on which the evaluation process is carried out either in its referential or in its pragmatic dimension. It is worth noting, however, that the two axes are independent of each other, and the morphological evaluation clearly shows that one should not attempt to match the values of the poles located on the same side of the axis. On the contrary, the correspondences are rather crossed: generally, the diminutive meanings of the referential pole (lower part of the RP axis) are rather positively connoted (upper part of the PP axis; e.g., *menotte* 'little hand, handy', from *main* 'hand') and conversely, the augmentative meanings (upper part of the RP axis) are rather negatively connoted (lower part of the PP axis; e.g., in Serbian, *sovuljaga* 'big, awful owl', from *sova* 'owl'). These correspondences, however, are only trends (see Grandi 2002: 34; see also Radial categories from Jurafsky 1996 or Mutz 2015): it is quite possible to find diminutives with negative connotations

(*réformette,* already cited, or *Paris* → *parigot* 'parisian', with a negative connotation) and augmentatives with positive connotations (e.g., in Serbian, *momčina* 'big guy', from *momak* 'boy'). A noun can also receive one connotation or the other depending on the context. For example, in French *églisette* could be interpreted as a 'small church (DOWN) that is quite charming (UP)', or a 'small church (DOWN) that is of little importance (DOWN)'[6].

When the Referential Pole alone is involved, which is quite infrequent, the derivative expresses one of the quantitative meanings (e.g., SMALL/BIG, FEW/MUCH) noted on the diagram:

(7)   dimin. *île* 'island' → *îlot* 'small island', *sauter* 'to jump' → *sautiller* 'to hop',
        pauvre 'poor' → *pauvret* 'fairly poor, poorish'

When only the Pragmatic Pole is involved, the derivative expresses one of the qualitative meanings (e.g., GOOD/BAD, EMPATHY/ANTIPATHY) noted on the diagram:

(8)   *frère* 'brother' → *frérot* 'bro', *traîner* 'to hang around' → *traînasser* 'to dawdle',
        merder 'to screw up' → *merdouiller* 'to mess things up'

In examples (8), the discrepancy from the standard does not concern the reduction of one or more referential dimensions expressed by the semantics of the base – *frère* and *frérot*, *traîner* and *traînasser*, etc. have exactly the same denotation –, but it concerns the standard associated with the use of the language itself. Thus, the terms cited under (8) can only be used in a colloquial or informal register. Other values are often present: a hypocoristic value (*frérot*) or a depreciative one (*traînasser*) for example.

In most cases, however, both poles are involved:

(9)   *bec* 'beak' → *bécot* lit. 'small beak', 'peck', *simple* 'simple' → *simplet* 'simple-
        minded', *beige* 'beige' → *beigeasse* 'loosely beige', *travailler* 'to work' → *travailloter*
        'to potter'

The evaluative derivatives[7] under (9) express both types of non-conformity with the standard: (a) they are pragmatically marked in that they do not belong to the standard

---

[6] Thanks to one of our reviewers for having led us to clarify this point through his/her remarks.

[7] Evaluative morphology is considered as part of derivational morphology (cf. in this respect the synthesis of Grandi 2017, and, among others, Dressler & Merlini Barbaresi 1994; Haspelmath 2010; Körtvélyessy 2015), contrary to Scalise (1984) who makes it a "third morphology", alongside inflectional and derivational morphology (see also Fortin 2011; Martín Calvo 2022). We will therefore speak of "evaluative derivatives", even

vocabulary but to distinctive registers, involving in many cases the speaker's positive or negative attitude; (b) they also express referential discrepancy from the standard: a *bécot* is a 'small kiss', i.e., a rather quick and light kiss, and *travailloter* is to work in a certain way: slowly or sluggishly.

3.2 The constraints leading to an approximative reading

Amiot & Stosic (2022) report that only a subset of diminutive derivatives can express this approximative meaning. To do this, they must display the following properties:

(i) Formally, homocategoriality must be at work, i.e., the category of the derivative must be identical to that of the base. This is the case of the nouns, adjectives and verbs in (10), where the input and the output of the derivation process belong to the same lexical category, and which all display an approximative meaning:

(10)   N/N: *livre*$_N$ 'book' → *livret*$_N$ 'booklet', *île*$_N$ 'island' → *îlot*$_N$ 'small island, islet'

           Adj/Adj: *gentil*$_A$ 'kind' → *gentillet*$_A$ 'kind enough', *jeune*$_A$ 'young' → *jeunot*$_A$ 'a bit young, immature'

           V/V: *voler*$_V$ 'to fly' → *voleter*$_V$ 'to flutter around', *boiter*$_V$ 'to limp' → *boitiller*$_V$ 'to hobble'

These examples show that the use of evaluation, as it reports a sort of discrepancy with respect to a standard, is closely linked to the difficulty of categorization (X is almost a Y but cannot be unconditionally categorized as a Y). This difficulty of categorization, which gives rise to the approximate interpretation, seems to imply a metalinguistic judgment: since the base name is not suitable to denote the entity in question, the diminutive suffixation makes it possible to create another term, which signals both similarity and non-conformity with the base. The homocategoriality can thus be thought of as a direct consequence of both the similarity of the referent to the standard and the impossibility of classifying it in another category. As a result, strictly speaking, only one referential category is cognitively processed, which implies, at the linguistic level, that the base and the derivative belong to the same lexical category. When two different referential categories are involved, the derivative cannot have this approximation meaning, as in examples (11):

---

for non-lexicalized forms, or "approximative derivatives", to refer to suffixed forms manifesting, respectively, an evaluative and/or an approximative meaning.

(11)   V/N: *sucer* 'to suck' → *suc**ette*** 'lollipop'

                  A/N: *bleu* 'blue' → *bleu**et*** 'cornflower', 'blueberry'

Sucette 'lollipop' and *bleuet* 'cornflower'/'blueberry' do not have an approximative meaning, even though they still appear to have an evaluative meaning[8].

However, homocategoricality is not a sufficient criterion as some homocategorical derivatives do not have this approximative meaning either:

(12)   N/N: *lune*$_N$ 'moon' → *lun**ette***$_N$ 'glasses'

                  N/N: *fenouil*$_N$ 'fennel' → *fenouill**et***$_N$ 'variety of apple that smells like fennel; fenouillet'

Although the nouns in (12) are homocategorical (they are denominal nouns), the derivatives do not express approximation because the construction of their meaning involves either a metaphor (glasses are shaped like small moons) or a metonymy (the smell of the apple reminds one of fennel), which leads to the processing of two distinct conceptual categories. There is no referential evaluation here (for example *lunette* does not denote a 'small moon'), nor, by way of consequence, any difficulty in categorization[9].

(ii) In addition, the evaluation made in reference to a standard, as mentioned above, must necessarily involve the Referential Pole, either exclusively, as in (13a), where the derivatives express only diminution (a *cuisinette* for example has the bulk of the properties of a cuisine 'kitchen', but it is smaller and less functional than a "normal" kitchen); or in association with the Pragmatic Pole, which is illustrated by the words in (13b) which, in addition to their meaning of diminution that allows an approximative interpretation, belong to non-standard language registers (hypocoristic, depreciative, colloquial). Only examples (13c), which exclusively involve the Pragmatic Pole, cannot express approximation:

---

[8] The evaluative meaning of words like *sucette* and *bleuet* is not obvious. Some authors have shown that, even if they are lexicalized, words suffixed with *-et* still denote small, manageable entities, etc. (see for example Fradin 2003; Meunier et al. 2007). Amiot & Stosic (2022) also show that many *-et* suffixed words are vernacular names of entities that often have scientific names. *Bleuet*, which is one of the vernacular names for *Centaurea cyanus*, is a good example of such a case. We are therefore faced with a pragmatic evaluation, not a referential one.

[9] This case is not very frequent in French, whereas it is widespread in Serbian, which is why, in Section 4.2, two types of homocategoriality are distinguished: a morpho-syntactic homocategoriality and a conceptual one. In French, the two types of homocategoriality generally coincide, except in cases such as this one, where only morpho-syntactic homocategoriality is present (about this distinction, see Section 4.2).

in this case the derivative has exactly the same denotation as its base, and only the register and the pragmatic value change.

(13) a. *cuisinette* (< *cuisine* 'kitchen') 'small kitchen, kitchenette'[10], *mordiller* (< *mordre* 'to bite') 'to nibble'

   b. *angelot* (< *ange* 'angel') lit. 'small angel', 'cherub', *suret* (< *sur* 'sour') 'tart, slightly sour', *amourette* (< *amour* 'love') 'casual love affair', *vivoter* (< *vivre* 'to live') 'to get by', 'to barely survive'

   c. *frérot* (< *frère* 'brother') hypocoristic 'bro', *traînasser* (< *traîner* 'to dawdle') depreciative 'to dawdle', *merdouiller* (< *merder* 'to screw up') informal 'to screw up'

3.3 Semi-categorization

Approximative derivatives cannot refer to autonomous conceptual categories, so they cannot integrate lexical hierarchies based on the hyperonymy/hyponymy relation (see Amiot & Stosic 2022). For example, unlike such compounds as *vert pomme* 'apple green', *vert émeraude* 'emerald green' or *vert d'eau* 'aqua green', the French adjective *verdâtre* 'greenish' cannot be considered a hyponym of *vert* 'green'. *Verdâtre* (< *vert* 'green') does not denote a kind of green but only a non-prototypical green, with a depreciative nuance (cf. **le verdâtre est une nuance de vert* 'greenish is a hue of green' vs. *le vert pomme est une nuance de vert* 'apple green is a hue of green'), which is usually the case with the suffix *-âtre*.

If these derivatives cannot integrate a hierarchy based on hypo-/hyperonymy relations, it is because it is not possible to identify a set of stable and recurrent properties – the *differentiae specificae* –, which would make it possible to constitute subclasses and be located among possible co-hyponyms (cf. *Les siamois et autres chats...* 'Siamese and other cats...' vs. **Les chatons et autres chats...* 'Kittens and other cats...'). For these reasons and based on a sort of conceptual derivation, what such evaluatives denote seems to correspond to kinds of "parasitic semi-categories" (see Amiot & Stosic 2022) rather than to true sub-categories of the category expressed by their base[11]. Consequently, it is difficult to establish the

---

[10] The term *kitchenette* is also used in French.

[11] Amiot & Stosic's term "parasitic semi-category" stems directly from Ungerer & Schmid's (1996: 74–76) notion of "parasitic categorization" that is used to explain a specific way of defining some superordinate categories without common "underlying gestalt that applies to all category members" (e.g., FURNITURE, FRUIT). To denote such categories, one can "'borrow' the gestalt properties of the superordinate category from the basic-level categories involved". See also Mihatsch (2007: 361) who argues for an extension of this

corresponding prototype, because their representation seems to be both highly dependent on the category fixed by the base, and not sufficiently elaborated in speakers' minds to give rise to an autonomous category. Such referents can be considered as gravitating in a kind of in-between space delimited by the core of the base category on the one hand, and by neighboring categories on the other hand.

As for non-approximative diminutives (e.g., *sucette* 'lollipop'), they generally denote entities that constitute autonomous categories, and thus appear in hierarchies based on the hyponymy/hyperonymy relation: a *sucette* 'lollipop' for example is a type of candy, just like a hard candy, an acid drop, a gummy worm, and so on. *Bonbon* 'candy' is therefore the hyperonym of *sucette* 'lollipop', which in return is its hyponym. And *sucette* 'lollipop' has other candy names as its co-hyponyms[12].

We now turn to suffixal augmentation, to see if it can also express approximation. As announced in the introduction, the data will be provided by Serbian.

4. Suffixal augmentation in Serbian

We assume that suffixal augmentation is construed according to the same principles as diminution: the evaluation is done through the same bipolar axis, but it targets the opposite pole and expresses an excess in relation to the norm (see Fig. 1 above).

This is very clear from the examples from Serbian below:

(14) a. *hrid* 'rock' → *hridina* 'big rock'

b. *dete* 'child' → *detište* 'pest child'

c. *kapetan* 'captain' → *kapetančina* 'famous, very worthy, admirable captain'

d. *magla* 'fog' → *magluština* 'very thick and awful fog'

e. *momak* 'boy' → *momčina* 'big good boy' 'big guy'

---

principle to the transfer of attributes from basic-level categories to categories belonging to other levels (including the lower one).

[12] An exception, however, is the evaluative derivatives which only involve the pragmatic pole (e.g., *frère* 'brother' → *frérot* 'bro'); these words cannot integrate a hierarchy of this type because they have exactly the same denotation as their base. There is no specific difference between them and their base and they cannot have the status of a hyponym.

4.1 Data collection

The study is based on a large sample of evaluative lexemes, namely 814 nouns, formed by a set of nominal suffixes traditionally considered as construing augmentatives (cf. Klajn 2003; Jovanović 2010; Bońkowski 2017).

Two main sources were used for data collection: the *Serbian Electronic Dictionary* (about 300 000 entries, see Simić 2005), and a 477-million-word Serbian corpus srWaC (v. 1.2), available in Sketch Engine[13], which are two complementary resources. By using various sampling methods, we obtained a sample of about 3300 candidates extracted from the dictionary and about 5000 candidates taken from srWaC. Only one tenth of the forms, namely 814 items, were selected as expressing evaluation[14]. The list of suffixes and the distribution of the 814 nouns collected for the purposes of this study are shown in Tab. 1.

Tab. 1: Distribution of nouns by suffix

Suffix	Nb occurrences	Suffix	Nb occurrences
-ina[15]	519	*-enda*	22
-ura	60	*-ište*	20
-ulja	51	*-sk(e/a)ra*	16
-uša	40	*-aga*	9
-č(u/a)ga	26	*-injara*	1
-uga	25	*-udža*	1
-(u/e)rda	24		

Some derivatives that we collected are present only in the dictionary (71,5%), others were found only in srWac (9,7%), and some appear in both resources (18.8%).

The bases these suffixes operate on were examined with reference to several criteria. First of all, we observed the part of speech (POS) of all the bases. Homocategoriality is at

---

[13] The *Serbian web corpus* (srWaC) is a Serbian corpus made up of texts collected from the Internet by Tomaž Erjavec and Nikola Ljubešić. The corpus was created in January 2014 (see Ljubešić & Klubička 2014).

[14] This score results from a great versatility of some of these suffixes. For instance, the ending *-ina* can form quality nouns from adjectives (e.g., *brz* 'fast' → *brzina* 'speed', *prazan* 'empty' → *praznina* 'emptiness'), nouns denoting meat from the corresponding nouns for animals (e.g., *prase* 'pig' → *prasetina* 'pork'), fraction nouns derived from numerals (e.g., *pet* 'five' → *petina* 'fifth'), place nouns derived either from verbs (e.g., *padati* 'to fall' → *padina* 'slope') or from nouns (e.g., *dô/dol* 'dell' → *dolina* 'valley'), or a series of nouns formed from adverbs (e.g., *mnogo* 'a lot of' → *množina* 'plural', *koliko* 'how much' → *količina* 'quantity') (see Klajn 2003: 95–98).

[15] The augmentative suffix *-ina* displays a range of variants: *-čina, -etina, -ština, -urina, -ušina, -uština*, and so on. Given the methodological approach of this study aiming at building a representative corpus of nominal augmentatives in Serbian, they were all grouped together in this study. It is clear, however, that the allomorphy of *-ina* deserves an in-depth empirical study (see also Klajn 2003: 101–104).

work in 84% of cases (15), but Serbian also allows the formation of nouns with augmentative meanings from verbs (16) and adjectives (17), in about 8% of cases for each[16]:

(15) a. *glava* 'head' → *glavurda* 'big and/or awful head'

b. *jež* 'hedgehog' → *ježina* 'big hedgehog'

c. *muškarac* 'man' → *muškarčina* 'big, tough and/or macho man'

(16) a. *nemati* 'to do not have' → *nemaština* 'scarcity'

b. *obiti* 'to burgle' → *obijačina* 'big and/or memorable burglary'

c. *pljuvati* 'to spit/speak ill of' → *pljuvačina* 'denigration/defamation'

d. *škrabati* 'to scrawl' → *škrabotina* 'scribble/daubing'

(17) a. *dobar* 'good' → *dobričina* 'a very kind person'

b. *glup* 'stupid' → *gluperda* 'a ridiculously stupid person'

c. *lenj* 'lazy' → *lenčuga* 'lazybones'

Second, we took into account several semantic features of the nouns serving as a base for forming augmentatives in order to better understand what kinds of concepts are best suited for such an interpretation. One of them is the category of animacy. Our data indicates that in this type of evaluation the base can refer to both animate and inanimate entities in more or less equal proportions (respectively 47,5% and 52,5%). The second one is the concrete-abstract distinction, which shows a clear imbalance in favor of concreteness: 97% of bases represent concrete denotata such as humans (e.g., *đak* 'pupil' → *đačina* 'big and/or very good pupil'), objects (e.g., *čizma* 'boot' → *čizmetina* 'big, awful boot'), animals (e.g., *ptica* 'bird' → *ptičurina* 'big and/or dangerous and/or disgusting bird'), and so on. We found only two abstract categories supporting augmentative evaluation: events (e.g., *bankrot* 'bankruptcy' → *bankročina* 'serious bankruptcy') and qualities (e.g., *bezobrazn*(ost) 'insolence' → *bezobraština* 'brazen insolence').

Another relevant feature of the data collected pertains to the way the opposition between the Referential and Pragmatic Poles (RP vs. PP) is exploited in augmentative

---

[16] One of our reviewers suggests that, in this case, the suffix *-ard* in French can also be considered as an augmentative suffix (*pleurnicher* 'to grizzle' → *pleurnichard* 'whiny', *gueuler* 'to shout' → *gueulard* 'loudmouth'), but in our opinion, these nouns rather express habituality, which "presupposes a more or less regular iteration of an event, such that the resulting habit is regarded as a characterizing property of a given referent. The notion of habituality is thus strictly related to iterativity, although the two should not be confused" (Lenci 2012).

suffixation in Serbian. Our corpus brings to the fore a very interesting interplay between them (see Tab. 2). First, one can observe that in 76% of cases, the two poles go hand in hand. Next, the "bigness meaning" most frequently has a pejorative value (71%) (e.g., *knjižurina* 'big and/or unpleasant book'), but some derivatives (about 4%) combine bigness and melioration (e.g., *herojčina* 'famous, admirable hero'). Note also that in a very few cases (9 nouns, 1%), the same augmentative item can, depending on its use, express either a positive or a negative attitude of the speaker (e.g., *mudarčina* 'wise man, sage' or 'someone who pretends to know everything'). Finally, in only 6% of cases, augmentation is not associated with any pragmatic reading (e.g., *koferčina* 'very big suitcase').

Tab. 2: Interplay between Referential and Pragmatic Poles

PRAG. POLE REF. POLE	PEJ	MEL	PEJ / MEL	Ø
BIG/MUCH/MANY/TOO MUCH	71%	4%	1%	6%
Ø	17%	1%	x	x

The same table also shows that in 18% of cases, the evaluative derivatives at study only involve either pejoration (17%) or melioration (1%), without affecting in any way the referential dimension. Putting aside the two least productive suffixes (-*injara* and -*udža*), the most prominent suffixes with this behavior are -*uša, -ura* and -*ulja*. As shown in Fig. 2, -*uša* is highly specialized in the expression of speaker affect, and -*ura* and -*ulja* do so in more than 50% of their derivatives in our corpus.
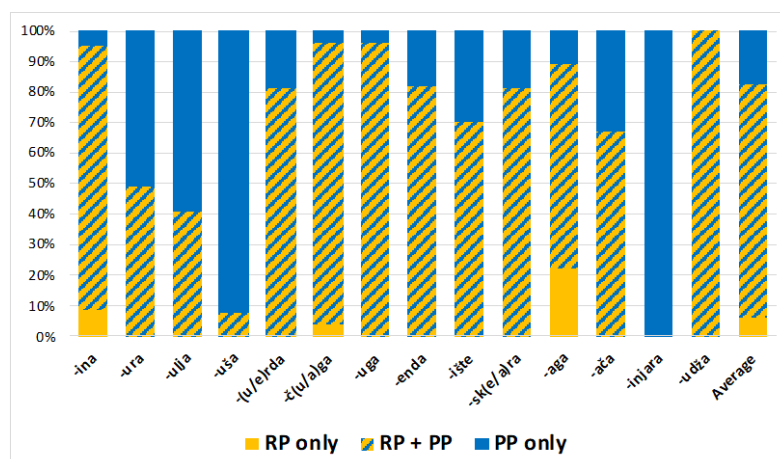
Fig. 2: Distribution of referential (RP) and pragmatic (PP) interpretations by suffix

Our sample thus provides a wide variety of nominal derivatives obtained by a series of so-called augmentative suffixes in Serbian (see Klajn 2003; Jovanović 2010). As such, it offers

a solid base for analyzing semantic values involved in this type of evaluative morphology, and consequently the relation between augmentation and approximation.

4.2 Augmentation and approximation: two irreconcilable values

As recalled above (see Section 3.2), only a subset of diminutive suffixed derivatives can express approximation. The approximative meaning is thus driven by two main constraints: homocategoriality and involvement of the referential dimension in the evaluation. In this section, we examine the relevance of these two constraints for an approximative reading of augmentative suffixed derivatives in Serbian.

The Serbian data suggest that two types of homocategoriality should be distinguished: a morpho-syntactic one and a conceptual one. The first type is illustrated by the examples in both (18) and (19), the latter only by the pairs in (18).

(18) a. N/N: *krv* 'blood' → *krvuština* 'very impure, sullied blood'

   b. N/N: *kuća* 'house' → *kućerina* 'big and/or creepy house'

   c. N/N: *rupa* 'hole' → *rupčaga* 'very big and/or very embarrassing hole'

(19) a. N/N: *krv* 'blood' → *krvulja* 'very nasty, perverse woman'

   b. N/N: *krokodil* 'crocodile' → *krokodiluša* 'very sly, treacherous, perfidious woman'

There is morpho-syntactic homocategoriality in all these examples because in all of them, the evaluative noun is formed using a nominal basis (N/N). But in (18a), we are dealing with only one conceptual category, the category of BLOOD[17], whereas the examples in (19) evoke two distinct conceptual categories, namely BLOOD and WOMAN in (19a), and CROCODILE and WOMAN in (19b). Consequently, what distinguishes the two sets of examples is the fact that in (17) the derivatives do not refer to autonomous conceptual categories, but rather to "parasitic" semi-categories, just as diminutives do (see Section 3.3 above), whereas in (19), even though both the bases and the derivatives belong to the category of nouns, and share the same root, two distinct conceptual categories are dealt with. Interestingly, in such derivatives, the so-called conceptual and evaluative morphology seem to be comfortably entwined, in that the evaluative derivatives perfectly fix new conceptual categories

---

[17] We are using small caps for conceptual categories.

that can integrate lexical (rather non-conventional) hierarchies. Nevertheless, these nouns have both an augmentative and a depreciative meaning, as shown by the glosses.

Following what has been shown for the approximative reading of diminutives (Section 3.2), examples such those in (18), involving both morpho-syntactic and conceptual homo-categoriality, are the best candidates for an approximative interpretation. However, this type of reading is not possible, whatever the semantic type of the base (see (18) above). The following examples stress the difference between diminution and augmentation with respect to approximation:

(20) a. N/N: *knjiga* 'book' → *knjižurak* 'booklet, small (cute) book' or 'book of poor value'

    b. N/N: *knjiga* 'book' → *knjižurina* 'big and/or unpleasant book'

(21) a. N/N: *gomila* 'pile' → *gomilica* 'little pile'

    b. N/N: *gomila* 'pile' → *gomiletina* 'big and/or embarrassing pile'

Even though the four derivatives undergo an evaluation process which involves deviation from the prototypical representation associated with the base, only the two derivatives in (20a) and (21a) involve approximation due to the marking of diminution, which suggests that the expected representation conveyed by the base is not reached. On the contrary, the augmentatives in (20b) and (21b) entail that the prototypical representation provided by their base is reached and even exceeded. By the very fact of excessiveness, they indicate that their referents are not prototypical members of the categories of books and piles, but this does not give them an approximative meaning. Consequently, in the case of augmentatives, homocategoriality does not trigger an approximative reading.

The involvement of the referential dimension in the process of evaluation is the second facilitating factor for an approximative reading in the case of diminutives. Consequently, one can expect that an approximative interpretation will be available for those augmentatives that involve the Referential Pole. This implies the modification of certain referential dimensions of the entities they denote, which is the case of a large part of our corpus. Again, this prediction does not work, as shown in (22). Although the evaluation with respect to a standard clearly entails the Referential Pole (it is about a big bitch, or a great robber), the approximative meaning is not present: *kurvetina* refers not to a usual whore but to an exceptional one, who carries out too many sexual acts, while *pivčuga* is used to

glorify the excellence of a given beer, which not only fulfils the requisite conditions to be categorized as beer, but largely exceeds them by its qualities.

(22) a. N/N: *kurva* 'whore, bitch' → *kurvetina* 'someone who engages in a lot of sexual acts for money, and/or big, fat bitch'

b. N/N: *lopov* 'robber' → *lopovčina* 'who steals a lot, impressive and/or disgusting robber'

c. N/N: *pivo* 'beer' → *pivčuga* 'true, genuine, great beer'

In all these examples, the evaluative process consists in picking out a given attribute or a set of attributes of the referent of the base, and in intensifying it along a scale in an ascending order from the norm[18].

In view of the above explanation, neither diminutives nor augmentatives can be considered as prototypical means for labelling stand-alone categories; in the vast majority of cases, they serve to label a kind of "semi-category", which is ontologically dependent on the category denoted by the base, but unable to form a separate category. However, our analysis clearly suggests that while diminutive suffixes can be considered to be, under certain conditions, approximators, i.e., triggers for approximation, augmentative suffixes cannot. In the following section, we explain why this is so.

5. Discussion

Taken together, our results allow greater insight into (i) the relationship between morphological augmentation and approximation (Section 5.1), (ii) the gradual nature of approximation (Section 5.2), and (iii) the peculiarities of approximation when conveyed by diminutive suffixation (Section 5.3). In the following sections, we discuss these three questions.

5.1 Why do only some deviations from the norm lead to approximation?

Since the formation of diminutive and augmentative meanings in the nominal domain involves the same principles (both types of evaluatives express non-conformity with respect

---

[18] Frequently, what is intensified is a salient bad or good property of the base itself; almost half of the bases in our corpus (46%) carry a negative connotation, 5% carry a positive one, and 46% of them are neutral. Such characteristics of the base are often crucial for the pragmatic value conveyed by evaluative derivatives: negatively connotated bases regularly lead to pejorative meanings (see 22a, b), positively connotated bases generally give rise to meliorative ones (e.g., *poštenjak* 'honest person → *poštenjačina* 'very honest and laudable person').

to the norm represented by the base, and constitute semi-categories that cannot be integrated into taxonomic hierarchies), it may be difficult to understand why diminutive meanings can, under certain conditions, express approximation while augmentative meanings cannot, regardless of the conditions considered ("full" homocategoriality and implication of the RP).

The only real difference between diminutive and augmentative evaluation lies in the result of the evaluation itself: the meaning of diminution appears only when, despite a very high similarity between the comparator (the base) and the compared (the derivative), the norm is not reached, whereas the meaning of augmentation appears when the norm is not only reached but also exceeded. Approximation thus seems to be necessarily associated with a certain deficiency and negativity. What is interesting is that this negativity does not seem specific to diminutive suffixation. Discontinuous reduplication in Italian studied by Masini & Di Donato (2023) is another very good example: in this process, the structure of the construction itself, [X *non* X], manifests in an almost iconic way both similarity and dissimilarity, and thus the underlying negativity, which can of course take many forms. This is also shown in the study of many of the processes addressed in this volume and is very well summarized in the paper by Hüning & Schlücker (2023: 106) who state that: "It is essential, however, that approximative formations always express that category X – however great the similarity – is ultimately not present, which is why approximation is also referred to as privative."

All this raises the question of whether approximative diminutives denote entities that are, or are not, included in the category denoted by the base.

5.2 Internal vs. external approximation: a solution?

Brucale & Mocciaro (2023), who analyze the degrees of approximation expressed by the Sicilian suffixes *-ḍḍu/ḍḍa*, propose an interesting distinction between internal and external approximation. Internal approximation "modifies the semantics of the base, but does not alter [its] categorial status" (Brucale & Mocciaro 2023: 28); this sort of approximation evaluates the proximity of the category to the center, i.e., to the prototype, and deals with peripheral members of a category (e.g., Sicilan: *sciccarieḍḍu* 'little donkey' < *sceccu* 'donkey'). External approximation on the other hand "approximates the category" and "does not

preserve the referential features of the base" (e.g., Sicilan: *fuochiḍḍi* 'fireworks' < *focu* 'fire'). It concerns referents that resemble the category but their actual belonging to it is not always clear-cut. The authors summarize this distinction in a figure:
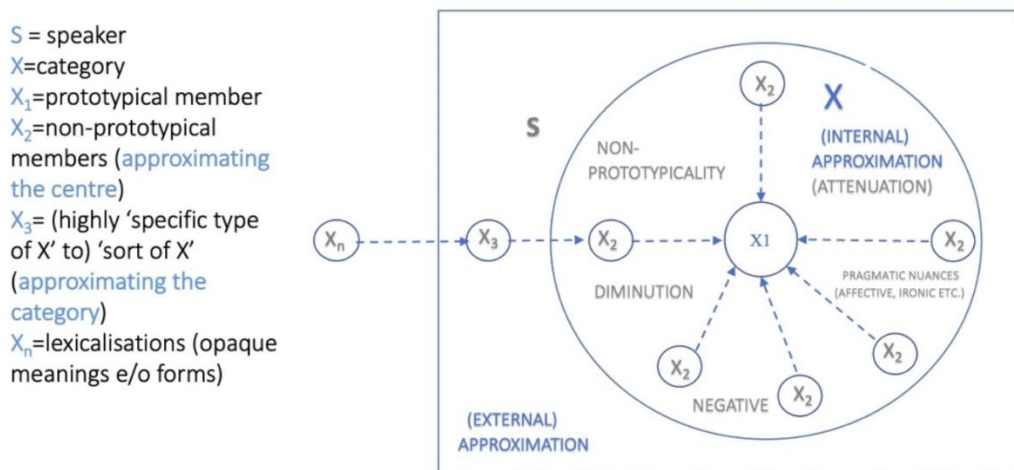
Fig. 3: Degrees of approximation: internal vs. external approximation
     (based on Brucale & Mocciaro 2023)

The distinction between the two types of approximation can be easily transposed to the approximative diminutives that we have studied. Generally speaking, as we have seen, they express non-conformity with respect to a norm, or a prototype, and as such they can be considered as "X_2", i.e., members falling under internal approximation, and more precisely diminution.

The use of certain diminutives sometimes even seems to call this peripheral categorial membership into question:

(23)　a.　*Pouvons-nous appeler cela une réforme? Une réformette plutôt!*
          'Can we call it a reform? A cosmetic reform rather!'

　　　b.　*Vraie réforme ou... réformette?* (*Le Parisien*)
          'Real reform or cosmetic reform?'

In (23a), the speaker's metalinguistic questioning leads to the opposition between *réforme* (the base) and *réformette* (the derivative). The use of the hedge *vrai* 'true' in (23b) leads to the same opposition. In these cases, the derivative seems to acquire greater autonomy with respect to the concept of the base, without being completely detached from it. In this type of use, the diminutive can switch into external approximation, and acquire the X_3 status (see Fig. 3), which often happens when the speaker's subjectivity comes into play.

This representation, however, cannot be transposed to augmentatives, which express non-prototypicality but not approximation.

5.3 Specificities of diminutive approximation

What seems fundamental in diminutive/augmentative suffixation is the fact that the evaluation is "measurative", i.e., it is processed in some way along the polar opposition UP-DOWN by means of the orientational metaphor as described by Lakoff & Johnson (1980); the norm, or the prototype, plays the role of zero point of the system (see Fig. 1). This is what induces, for diminution, the possibility of expressing approximation (the norm is not reached), and for augmentation, the impossibility of expressing it (the norm is not only reached, it is exceeded).

We argue that this measurability distinguishes diminutive approximation from other types of approximation, e.g., from the suffixal approximation expressed in the nouns and adjectives in *-oïde* and *-forme* mentioned at the beginning of the article. Here are two nominal examples[19]:

(24)   a.   *organoïde* 'organoid': 'complex collections of cells that bear a resemblance to human tissue'
       b.   *insectoïde* 'insectoid': 'species with insect-like features'

In both examples, the derivative qualifies an entity that only partially possesses the properties associated with the base: organoids are not organs, but clusters of stem cells created to have the same properties as real organs in order to replace them in case of disease. As for insectoids, in science fiction or ufology they are often robots, evil or not, whose form imitates that of insects.

So, like diminutive/augmentative derivatives, nominal (and also adjectival) derivatives in *-oïde* express both resemblance and difference with respect to an entity that serves as a norm, represented by the base. However, the evaluation process does not involve either of the bipolar axes linked to the Referential Pole (e.g., SMALL/BIG, FEW/MUCH) or the Pragmatic Pole (e.g., BAD/GOOD, ANTIPATHY/EMPATHY): in French, *-oïde* derivatives do not express measurative meanings, either diminution or augmentation. Therefore, the approximation expressed by *-oïde* derivatives is not measurative.

---

[19] Although *-oïde* forms mostly adjectives, we have chosen to take nominal examples in order to facilitate the comparison, especially with Serbian augmentatives, which all belong to the noun category.

The paper on *pseudo-* in Greek and French by Vassiliadou et al. (2023) highlights, we think, the difference between approximation expressed by diminutive suffixation in the nominal domain and other modes of word formation. According to the authors:

> Pseudo-X always implies categorization; it presupposes a comparison between what it designates and an expression X, and it consists in an evaluation of the categorial adequacy between what pseudo-X designates and the category X itself. This evaluation is negatively polarized, i.e., it is focused on features of difference, or rather of deviation from the category (see the term of disproximation; Cappelle et al. 2023). Thus, the referent of pseudo-X deviates in one way or another from X. (Vassiliadou et al. 2023: 248)

As with diminutive suffixation, prefixation with *pseudo-* evaluates the overall degree of matching with a standard, so that this evaluation is done globally, without any specific bipolar semantic dimension such as that underlying evaluation as defined here (see Fig. 1 supra). The same is true for the negative pragmatic values that characterize most derivatives in *pseudo-*: in fact, *pseudo-*X never manifests a meaning that could be called "measurative" (more/less) but rather global "qualitative" meanings such as 'fake X' or 'imitation of X', etc. (see Vassiliadou et al. 2023 for a detailed analysis of the different possible meanings).

In a way, measurability could be a criterion to distinguish two sorts of approximation, i.e., non-measurative approximation, which is probably the most widespread as conveyed by *pseudo-*, or *-oïde*, and the measurative approximation, which is restricted, as far as we know, to nominal diminutive suffixation, and probably to the different categories of diminutive suffixation.

This distinction between measurative and non-measurative approximation could be related to the distinction that Amaral & Del Prete (2010) make between scalar and non-scalar approximation in their analysis of the adverb *quasi* 'almost' in Italian. However, the theoretical approach (fundamentally logical) and the type of data studied (the adverb *quasi* as a modifier of temporal connectives) are not directly transposable to our analyses and the comparison would require further investigation.

6. Concluding remarks: approximation between the norm and negation of the category

Like other papers in this volume, our study suggests that approximation hosts a great diversity of semantic values going from a slight discrepancy with respect to the norm or the prototype to calling into question the category membership. The distance between the latter and the prototypical members is very great, but not sufficient to block the possibility of their representation in terms of a given category. Evaluative morphology, and more precisely diminution, participates in providing means for denoting such peripheral members without completely negating their dependance on the category of the base. This leads to the blurring of the category boundaries.

Based on the schematic representation of the scope of approximation proposed in the introductory paper by Masini et al. (2023) and Brucale & Mocciaro (2023), we can summarize the way in which evaluative morphology contributes to the construction of approximation as depicted in Fig. 4:
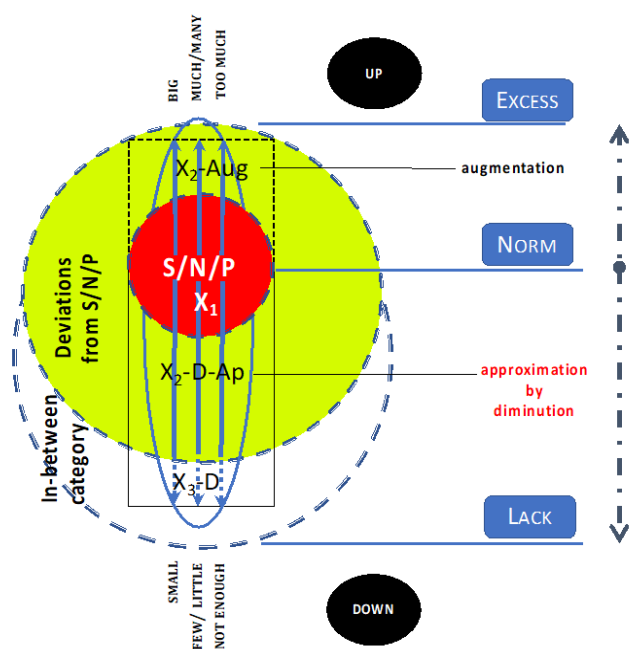
Fig. 4: The scope of approximation as conveyed by evaluative morphology, based on Fig. 1, Fig. 3 above, and the schema by Masini et al. (2023)

The red circle represents the norm or the standard (S/N/P, X_1), the green one a range of peripheral, more or less non-prototypical members (X_2), whereas the grey zone represents a kind of in-between category containing highly problematic, borderline members (X_3).

Due to its grounding in the UP/DOWN orientational metaphor and to a series of bipolar measurative meanings that it is based on, diminutive morphology occupies only a specific portion of the area of approximation, namely the blue rectangle, located in the lower part of the figure. This further means that approximation can only be expressed by a subset of diminutives, those for which the evaluation involves the referential pole, and which express a lack of a series of properties allowing a given referent to adequately integrate the category denoted by the base (X_2-D-Ap). That can be considered as a default interpretation. However, since the discrepancy with respect to the norm can be more or less marked, diminutive approximation can encroach on the grey area (X_3-D-Ap, see examples (23) supra) and come close to members generally described by devices with a privative meaning (e.g., *pseudo*-X, *simili*-X in some of their uses: *pseudo-cuir* 'pseudo-leather' or *simili-cuir* 'imitation leather' for example). This reading is generally prompted by the context. Conversely, as one might expect, augmentation does not have a place in the figure since it cannot express approximation: the data from Serbian clearly suggest that once the norm is reached, approximation necessarily disappears (X_2-Aug, without Ap(proximation)). Nevertheless, augmentatives, which express an excess compared to a norm, can convey non-prototypicality (X_2). That is what the empty rectangle in the upper part of the figure materializes. One can expect that these observations hold equally true for evaluative prefixes (*macro-, mega-, maxi-, archi-, hyper-, extra-, super-, ultra-, sur-*) in French, but this deserves an in-depth empirical study of these morphemes.

Contrasting diminutives and augmentatives in the light of different kinds of evaluative devices in French and Serbian has allowed us to highlight the specificity of diminutive approximation, which is fundamentally measurative, contrary to non-measurative approximation expressed by other morphological processes.

References

Amaral, Patricia & Fabio Del Prete. 2010. Approximating the limit: The interaction between quasi 'almost' and some temporal connectives in Italian. *Linguistics and Philosophy* 33(2). 51–115. DOI: 10.1007/s10988-010-9075-0.

Amiot, Dany & Dejan Stosic. 2011. *Sautiller, voleter, dansoter*: évaluation, pluriactionnalité, aspect. In Eugenia Arjoca-Ieremia, Cécile Avezard-Roger, Jan Goes, Estelle Moline & Adina Tihu

(eds.), *Temps, aspect et classes de mots: études théoriques et didactiques*, 277–297. Arras: Artois Presses Université.

Amiot, Dany & Dejan Stosic. 2019. Motion verbs and evaluative morphology. In Michel Aurnague & Dejan Stosic (eds.), *The Semantics of Dynamic Space in French: Descriptive, Experimental and Formal Studies on Motion Expression*, 180–215. Amsterdam: Benjamins.

Amiot, Dany & Dejan Stosic. 2022. Evaluative morphology: From evaluation to approximation and semi-categorization. In Hélène Vassiliadou & Marie Lammert (eds.), *A Crosslinguistic Perspective on Clear and Approximate Categorization*, 53–94. Newcastle upon Tyne: Cambridge Scholars Publishing.

Amiot, Dany & Edwige Dugas. 2020. Combining forms and affixoids in morphology. In Rochelle Lieber (ed.), *Oxford Encyclopedia of Morphology*. Oxford: Oxford University Press.

Bauer, Laurie. 1997. Evaluative morphology: In search of universals. *Studies in Language* 21(3). 533–575. DOI: 10.1075/sl.21.3.04bau.

Bertinetto, Pier Marco & Alessandro Lenci. 2012. Habituality, pluractionality, and imperfectivity. In Roberto Binnick (ed.), *The Oxford Handbook of Tense and Aspect*, 852–880. Oxford: Oxford University Press.

Bońkowski, Robert. 2017. Sufiksalne rzeczownikowe derywaty augmentatywne w serbskiej leksyce potocznej [Suffixable noun augmentative derivatives in Serbian common lexis]. *Zeszyty Cyrylo-Metodiańskie* 6. 57–68.

Brucale, Luisa & Egle Mocciaro. 2023. Approximation through suffixation: *-ḍḍu/-a* in Sicilian. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1). 27–51.

Cappelle, Bert, Robert Daugs & Stefan Hartmann. 2023. The English privative prefixes *near-*, *pseudo-* and *quasi-*: Approximation and 'disproximation'. *Zeitschrift für Wortbildung*/Journal of *Word Formation* 7(1). 52–75.

Channell, Joanna. 1994. *Vague Language*. Oxford: Oxford University Press.

Dressler, Wolfgang U. & Lavinia Merlini Barbaresi. 1994. *Morphopragmatics. Diminutives and Intensifiers in Italian, German, and Other Languages*. Berlin & New York: De Gruyter.

Efthymiou, Angeliki. 2017. Intensification and deintensification in Modern Greek verbs. *Lexis 10*. DOI: 10.4000/lexis.1089.

Fradin, Bernard. 2003. Le traitement de la suffixation en *-et*. *Langages* 152. 51–77. DOI: 10.3406/lgge.2003.2439.

Fradin, Bernard. & Fabio Montermini. 2009. La morphologie évaluative. In Bernard Fradin, Françoise Kerleroux and Marc Plénat (eds.), *Aperçus de morphologie du français*, 231–266. Vincennes: Presses Universitaires de Vincennes.

Fortin, Antonio. 2011. *The Morphology and Semantics of Expressive Affixes*. Doctoral dissertation, Oxford : Oxford University.

Gerhard-Krait, Francine & Hélène Vassiliadou. 2017. Clapotis, murmures et autres manifestations sonores: les méandres de l'approximation catégorielle. *Syntaxe et Sémantique* 18(1). 19–43. DOI: 10.3917/ss.018.0019.

Grandi, Nicola. 2002. *Morfologie in contatto. Le constrizioni valiative nelle lingue del Mediterraneano*. Milan: FrancoAngeli.

Grandi, Nicola & Lívia Körtvélyessy. 2015. Introduction: Why evaluative morphology? In Nicola Grandi & Lívia Körtvélyessy (eds.), *Edinburgh Handbook of Evaluative Morphology*, 3–20. Edinburgh: Edinburgh University Press. DOI: 10.3366/j.ctt1g09zkx.7.

Grandi, Nicola & Lívia Körtvélyessy (eds.). 2015. *Edinburgh Handbook of Evaluative Morphology*. Edinburgh: Edinburgh University Press.

Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in cross-linguistic studies. *Language* 86(3). 663–687. DOI: 10.1353/lan.2010.0021.

Hüning, Matthias & Barbara Schlücker. 2023. Approximation and comparison in word-formation. The case of denominal adjectives in Dutch, German, and English. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1). 101–129.

Jouitteau, Mélanie (ed.). 2009–2023. *ARBRES, wikigrammaire des dialectes du breton et centre de ressources pour son étude linguistique formelle*. IKER, CNRS. http://arbres.iker.cnrs.fr (accessed 13 December 2022).

Jovanović, Vladan. 2010. *Deminutivne i augmentativne imenice u srpskom Jeziku* [Diminutive and augmentative nouns in Serbian]. Beograd: Institut za srpski jezik SANU.

Jurafsky, Dan. 1996. Universal Tendencies in the Semantics of the Diminutive. *Language* 72(3). 533–578. DOI: 10.2307/416278.

Klajn, Ivan. 2003. *Tvorba reči u savremenom srpskom jeziku* [Word formation in contemporary Serbian], vol. 2. Belgrade: Institut za srpski jezik SANU, Matica srpska.

Kleiber, Georges. 1987. Quelques réflexions sur le vague dans les langues naturelles. In Sylvie Mellet (ed.), *Études de linguistique générale et de linguistique latine offertes en hommage à Guy Serbat*, 157–172. Paris: Société pour l'Information Grammaticale.

Körtvélyessy, Lívia. 2015. *Evaluative Morphology from a Cross-linguistic Perspective*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Lakoff, George. 1973. Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic* 2. 458–508. DOI: 10.1007/BF00262952.

Lakoff, George & Mark Johnson. 1980. *Metaphors We Live By*. Chicago: University of Chicago Press.

Ljubešić, Nikola & Filip Klubička. 2014. *{bs,hr,sr}WaC – Web corpora of Bosnian, Croatian and Serbian. Proceedings of the 9th Web as Corpus Workshop (WaC-9)*, 29–35. Gothenburg: Association for Computational Linguistics.

Martín Calvo, Rafael. 2022. *An Analytical Framework for the Description of Evaluative Morphology Resources: Contrastive Analysis of Evaluative Forms in Spanish and Latvian*. Doctoral dissertation, Ventspils: Ventspils University College.

Masini, Francesca & Jacopo Di Donato. 2023. Non-prototypicality by (discontinuous) reduplication: the N-*non*-N construction in Italian. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1). 130–155.

Masini, Francesca, Muriel Norde & Kristel Van Goethem. 2023. Approximation in morphology: A state of the art. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1). 1–26.

Mel'čuk, Igor. 1994. *Cours de morphologie générale (théorique et descriptive)*. Vol. 2. Montréal: Les Presses de l'Université de Montréal/Paris: CNRS Éditions.

Meunier, Fanny, Georgette Dal, Bernard Fradin & Nabil Hathout. 2007. Rôle de la productivité morphologique dans la reconnaissance des lexèmes complexes. In *Morphologie à Toulouse. Actes du colloque international de morphologie 4e Décembrettes*, 289–309. Munich: Lincom Europa.

Mihatsch, Wiltrud. 2007. Taxonomic and meronomic superordinates with nominal coding. In Andrea Schalley & Dietmar Zaefferer (eds.), *Ontolinguistics: How Ontological Status Shapes the Linguistic Coding of Concepts*, 359–377. Berlin: De Gruyter.

Mihatsch, Wiltrud. 2009. L'approximation entre sens et signification: un tour d'horizon. In Dominique Verbeken (ed.), *Entre sens et signification. Constitution du sens: points de vue sur l'Articulation sémantique-pragmatique*, 100–116. Paris: L'Harmattan.

Mutz, Katrin. 2015. Evaluative morphology in a diachronic perspective. *Edinburgh Handbook of Evaluative Morphology*, 142–154. Edinburgh: Edinburgh University Press.

Plénat, Marc. 1999. Distribution des suffixes évaluatifs en français. *Silexicales* 2. 179–188. https://stl.univ-lille.fr/fileadmin/user_upload/laboratoires/stl/Plenat_Reperes_morphologie_p114-124.pdf (accessed 18 May 2021).

Prince, Ellen F., Charles Bosk & Joel Frader. 1982. On hedging in Physician-Physician Discourse. In Robert J. Di Pietro (ed.), *Linguistics and the Professions*, 83–97. Norwood: Ablex.

Simić, Milorad. 2005. *Srpski elektronski rečnik* [Serbian electronic dictionary]. CD-ROM.

Scalise, Sergio 1984. *Generative morphology*. Dordrecht: Foris.

Stosic, Dejan & Dany Amiot. 2019. Motion verbs and evaluative morphology. In Michel Aurnague & Dejan Stosic (eds.), *The Semantics of Dynamic Space in French. Descriptive, Experimental and Formal Studies on Motion Expression*, 179–215. Amsterdam & Philadelphia: Benjamins.

Stump, Gregory. 1993. How peculiar is evaluative morphology? *Journal of Linguistics* 29. 1–36. DOI: 10.1017/S0022226700000037.

Ungerer, Friedrich & Hans-Jörg Schmid. 1996. *An Introduction to Cognitive Linguistics*. London: Longman.

Vassiliadou, Hélène, Francine Gerhard-Krait, Georgia Fotiadou & Marie Lammert. 2023. *Pseudo*(-) in French and Greek: categorization and approximation. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1). 234–262.

Vassiliadou, Hélène & Marie Lammert (eds.). 2022. *A Crosslinguistic Perspective on Clear and Approximate Categorization*. Cambridge: Cambridge Scholars Publishing.

Dejan Stosic

Université Toulouse-Jean-Jaurès / Centre national de la recherche scientifique (CNRS)

Laboratoire Cognition, Langues, Langage, Ergonomie (CLLE) – UMR 5263

F–31058 Toulouse cedex 9, France

dejan.stosic@univ-tlse2.fr

Dany Amiot

Université de Lille / Centre national de la recherche scientifique (CRNS)

Laboratoire Savoirs, Textes, Langage (STL) – UMR 8163

F–59000 Lille, France

dany.amiot@univ-lille.fr

Yvonne Treis

The approximative derivation in Kambaata (Cushitic)

Abstract: Kambaata, a Cushitic language of Ethiopia, has a productive, word-class changing approximative derivation. The process marked by the suffix *-lab* takes adjectival, nominal and verbal roots as inputs and generates approximatives that belong to the adjective word class, whose members are characterised by their agreement behaviour in the NP. The meaning of the Kambaata approximatives is dependent on the word class of its input. Deadjectival approximatives express a vague, attenuated degree of a property, deverbal forms convey similarity or near-identity in manner, while denominal forms express similarity in nature or function and create *ad hoc* categories. The sub-meanings are subsumable under an overarching meaning of approximation, understood as deviation from a prototypical realisation. The source of the *-lab* suffix is a verb 'resemble'. From a Cushitic perspective, the Kambaata approximative derivation is unique of its kind. No related language is known to have approximative morphology.

Keywords: approximation, derivation, adjective, adjectivisation, Cushitic

1. Introduction[*]

Kambaata, a Cushitic language spoken in Ethiopia, has a productive, recently grammaticalised approximative derivation. The derivational process marked by the suffix *-lab* (APRX) generates adjectives on the basis of adjectival (1), verbal (2) or nominal roots (3), in order to express a reduced (attenuated) degree of a property, e.g. 'red' > 'reddish' (1), a near prototypical way of carrying out an action, e.g. 'run' > '(in an) almost running (manner)' (2), or a resemblance in nature and character, e.g. 'needle' > 'needle-like (object)' (3).

---

(1) *biishsh-i-**láb**-aa*   *kin-íin*

    red-EP-**APRX**-M.OBL   stone-M.ICP

    'with reddish stones' (Saint-Exupéry 2018: 19)

(2) *dagud-**lab**-á*   *marámm* (...)

    run-**APRX**-M.ACC   walk.1SG.PFV.CVB

    '(I) went (there) almost running (...)'

(3) *marf-**lab**-á*   *ut-ichch-ú*

    needle-**APRX**-M.ACC   thorn-SGV-M.ACC

    'a needle-like thorn'

In contrast to some other Cushitic languages in which the existence of adjectives is debatable (see, for instance, Banti 1988; Mous 2012: 358, 377–379), Kambaata has a clearly identifiable adjective word class, which is defined on the basis of morphosyntactic criteria. Apart from many morphologically simple adjectives, Kambaata has several adjectivising processes that allow for the creation of a theoretically unlimited number of derived adjectives (see Alemu 2016, which contains not less than 1731 (sub-)entries labelled "sc", the language-internal abbreviation for "adjective").[1]

This paper focusses on a synchronic analysis of the Kambaata approximative derivation from a typological perspective. Earlier work on Kambaata adjectives has overlooked this derivational process, even though fairly detailed descriptions of the inflectional and derivational morphology of adjectives are found in Treis (2008: 254–302; 2011; 2017a: 349–351). The paper also proposes a likely diachronic source for the approximative morpheme and thus seeks to enrich the discussion of possible sources of approximative morphology cross-linguistically. Finally, this study wants to further our understanding of adjectival derivational processes in Cushitic – a field that is hitherto hardly explored (see the very brief notes in Mous 2012: 379; Shay 2014: 581).

The discussion is structured as follows. After a brief general introduction into the language and the data collection methodology in Section 2, I sketch the typological profile of the language and introduce the defining features of open word classes in Section 3. Section 4 gives an overview of Kambaata's adjectivising processes. Section 5, the core of the paper, discusses the morphology, syntax and semantics of approximative adjectives, shows that

---

[1] The label "sc" stands for *su'mm-caakkisaanchu* 'adjective'; an analysis of this term is later provided in (22).

an overarching meaning can be attributed to *-lab* and points out competing processes with an approximative meaning. Section 6 discusses the likely origin of the *-lab* derivation and proposes a grammaticalisation chain. Section 7 concludes the paper.

2. The language

2.1 Classification and speaker area

The Kambaata language (ISO-code 639-3: ktb, Glottolog code: kam1316) is spoken by the Kambaata, Xambaaro and Donga people, who settle around the Hambarrichcho massif in southern Ethiopia, about 300 km southwest of the Ethiopian capital Addis Ababa. According to the last census, Kambaata has more than 600,000 speakers (Central Statistical Agency 2007: 74), the large majority of which live in the Kambaata-Xambaaro Zone of the Southern Nations, Nationalities and Peoples' Regional State. Kambaata belongs to the Highland East Cushitic branch of the Cushitic family (Afroasiatic phylum). The immediate neighbours of the Kambaata are speakers of other Highland East Cushitic languages (Hadiyya [hdy] and Alaaba [alw]) and Ometo languages of the Omotic family (Wolaitta [wal] and Dawro [dwr]). Amharic [amh], the Ethiopian lingua franca, is the most important second language of Kambaata speakers.

2.2 Orthography

The official Kambaata orthography is based on the Roman script (Treis 2008: 73–80; Alemu 2016) and used with minimal modifications to transcribe the data in this article. The following graphemes are not in accordance with IPA conventions: <ph> /p'/, <x> /t'/, <q> /k'/, <j> /dʒ/, <c> /tʃ'/, <ch> /tʃ/, <sh> /ʃ/, <y> /j/ and <'> /ʔ/. Geminate consonants and long vowels are marked by doubling, e.g. <shsh> /ʃː/ and <ee> /eː/. Nasalisation is marked by a macron, e.g. <ā> /ã/. Consonant clusters consisting of a glottal stop and a simplex sonorant are spelt as trigraphs, e.g. <'rr> /ʔr/, to distinguish them from laryngealised sonorants, e.g. <'r> /r'/. The minor adaptations to the official orthography made in this contribution concern the additional marking of phonemic stress by an acute accent and the consistent notation of the phonemic glottal stop whenever it occurs in word-medial and word-final position.

2.3 Data collection

The data for this paper comes from a corpus of recorded narratives and conversations, my field notes of volunteered or elicited data as well as a corpus of locally published written texts. In the fieldwork data collected between 2002 and 2007 in the Kambaata-Xambaaro Zone, no derivates with *-lab* are attested – which explains why this process is not addressed in Treis (2008). The first two attestations are in a recording of a conversation made in 2016, which permitted the elicitation of some additional examples. The database on *-lab* derivates was significantly extended when Alemu's (2016) dictionary was published. The monolingual definitions that follow each lexical entry in this comprehensive work are a rich data source for the study of the approximative derivation; the definitions contain altogether as many as 61 types (144 tokens) of *-lab* derivates. In a fieldtrip in 2018, I extracted these examples from the dictionary, discussed their forms and meanings with two native speakers (dwd, swd) and asked them to exemplify their use in near-natural sentences or mock-dialogues, i.e. question-answer or statement-comment pairs of two imagined interlocutors. In the following years (2019–2022), 14 more *-lab* tokens were obtained, at different occasions, in solicited and unsolicited data from three different speakers (ts, dwd and ayz). Finally, the recently published collection of Kambaata idioms (Alamu 2022) contains 9 additional examples. At present, my database contains altogether 84 types of *-lab* derivates (209 tokens).

Three categories of data are distinguished by the labels after each example: (i) recorded, spontaneously produced data (marked by the initials of the recorded speaker's name, the date and the file number), (ii) written data (marked by a literature reference) and (iii) data elicited and volunteered in the field and in remote fieldwork sessions (labelled "elicited"). In the elicitation process, translation elicitation was intentionally avoided. Speakers were instead provided with word forms for which they formulated example sentences or dialogues, or they were asked to paraphrase or reformulate existing examples.

3. Typological profile and word classes

Kambaata has a head-final constituent order; the final element in a sentence is usually a fully finite main verb or a copula. The language is agglutinating-fusional and (almost)

exclusively suffixing. Apart from a number of closed classes, the following open word classes can be defined morphosyntactically: verbs, nouns, adjectives, ideophones and interjections. Verbal, nominal and adjectival roots are bound. Lexical roots and derivational morphemes are undetermined for stress, while inflectional morphemes consist of a segmental component and a suprasegment (stress) that is realised, dependent on the grammatical category to be marked, on a specific syllable of the inflected word form.

Kambaata is a nominative-accusative language. The nominative is the subject case; the accusative marks direct objects and certain adverbial constituents, and also serves as the citational form of nouns. Nouns are obligatorily marked for gender (masculine vs. feminine). The assignment of grammatical gender is mostly arbitrary, with the exception of nouns referring to human beings and higher animals, where it is sex-based; see, e.g., the arbitrary assignment of masculine gender to the term for a whole enset leaf,[2] *habar-á*, and of feminine gender to the term for an enset leaf from which the midrib was torn off and which is used as underlay, padding or wrapping, *faatt-áta* (Tab. 1). Nouns distinguish nine case forms, all of which are marked by a segmental suffix and a specific stress pattern. Stress is phonemic and serves – either alone or in combination with a segmental suffix – to distinguish between grammatical forms of one lexeme. Nouns are categorised into altogether 21 declensions (Treis 2008: 103).

Tab. 1: Case paradigm of a masculine noun (declension M1 in -*á*) and a feminine noun (declension F1a in -*áta*)

Case form	*habar-á* (M) 'enset leaf'	*faatt-áta* (F) 'enset leaf without the midrib'
Accusative	*habar-á*	*faatt-áta*
Nominative	*habár-u*	*fáatt-at*
Genitive	*habar-í*	*faatt-á*
Dative	*habar-íi(ha)*	*faatt-áa(ha ~ ta)*
Ablative	*habar-íichch*	*faatt-áachch*
Instrumental/comitative/perlative	*habar-íin*	*faatt-áan*
Locative	*habar-áan*	*faatt-áan*
Oblique/vocative	*habár-a*	*fáatt-a*
Predicative (with COP2)	*habár-a*	*fáatt-a*

---

[2] Enset (*Ensete ventricosum*) is a food plant cultivated in the highlands of southern Ethiopia.

Adnominal adjectives, numerals and demonstratives agree with their head noun in case and gender. The case system of adnominals is reduced to three forms: nominative, accusative and oblique. Adjectives fall into five declensions (Treis 2008: 256), of which one is exemplified in Table 2.

Tab. 2: Case/gender inflection of an adjectival modifier: The example of *farr-á(ta)* 'bad' (declension A1)

Case form	M	F
Accusative	*farr-á*	*farr-áta*
Nominative	*fárr-u*	*fárr-at*
Oblique	*fárr-a(a)*	*fárr-a(ta)*

The oblique form of adnominals signals agreement with a non-nominative, non-accusative head noun; see *tordúm-aa* before the instrumental head noun in (4).

(4)   *Bu'll-í*   *hogob-úta*   **tordúm-aa**   *birr-íin*
            enset_flour-M.GEN   donkey_load-F.ACC   ten-M.OBL   birr-M.ICP

            hir-éenno
            buy-3HON.IPFV

            '(In the old days,) one bought a donkey load of enset flour for (lit. with) 10 (Ethiopian) birr.' (EK2016-02-23_003)

An adjective (or a numeral) may also function as the head of an NP.[3] As such, it has the same case-marking potential as a noun and is marked for one of the nine nominal cases (Tab. 1); see, e.g., the adjective in the instrumental-comitative-perlative (ICP) case in (5) and the nominative and accusative adjectives in (6). As NP head, the adjective adopts the gender of the omitted head noun, which is masculine in (5) and (6).

(5)   **Hiil-íin**   *hoshsh-ó-o(hu)*   *hāy-íin*
            bad-M.ICP   pass_the_day-3M.PFV.REL-NMZ1.M.NOM   front_leg-M.ICP

            kad-áno
            kick-3M.IPFV

            [Proverb] 'The one who has passed the day with a bad one (i.e. in bad company) kicks with the front legs.' (Alamu & Alamaayyo 2017: 80)[4]

---

[3] In contrast to adjectives and numerals, demonstratives distinguish between adnominal and pronominal forms.

[4] Stress marks, glosses and translations of all data cited from local Kambaata publications are mine.

(6) *Kánn*   *wud-í*   **qoxar-á**   *hikkáaph*   *wud-í*
     A_DEM1.M.OBL   side-M.GEN   strong-M.ACC   A_DEM4.M.OBL   side-M.GEN

     qoxár-u   *uujj-ee'-íichch*   *ís*   *Leegáam-o-ot*
     strong-M.NOM   make_fall-3M.PFV.REL-ABL   3M.NOM   L.-M.OBL-COP3

     'If the strong one (i.e. strong wrestler) of that side (over there) takes down the strong one of this side, he is (called) "Leegaamo".' (EK2016-02-23_001)

The feminine adjective *híil-at* 'bad' in (7) refers to an implicit plural referent – note that two (of three) plurative markers in Kambaata are grammatically feminine (Treis 2014: 115).

(7) *Wól-it*   **híil-at**   *iill-ít*   *ul-tún-ka-'nne*
     other-F.NOM   bad-F.NOM   reach-3F.PFV.CVB   touch-3F.JUS-NEG3-2PL.OBJ

     [Blessing:] 'May other bad (things) not reach (and) affect you (PL)!' (AN2016-02-19_001)

Adjectival predicates agree with their subject in gender, are marked for the predicative (pred) case and combine with the ascriptive-identificational copula *-(h)a(a)* (M.COP2)/ *-ta(a)* (F.COP2), as in (8).[5]

(8) *Tí*   *hóol-ch-ut*   *abb-ís-s*
     A_DEM1.F.NOM   sheep-SGV-F.NOM   exceed-CAUS1-3F.PFV.CVB

     moos-áan-ch-u-ta
     disease-AG-SGV-F.PRED-F.COP2

     'This sheep (F) is very sickly.' (Saint-Exupéry 2018: 14)

As seen in the glosses of the preceding examples, verbs inflect for aspect, mood, polarity and dependency status. All verbs apart from verbal nouns carry simple or bipartite subject indexes. Direct and indirect objects can be expressed by suffixed pronouns at the right edge of the inflected verb (7).

---

[5] Numeral and demonstrative predicates take a different copula: -V*t* (COP3).

4. Overview of adjectival derivation and compounding

Kambaata has a large class of basic adjectival lexemes, such as *farr-á(ta)* 'bad' (Tab. 2). All simple, underived adjectival roots in the language have a corresponding inchoative-stative property verb of the same root, i.e. *farr*-Case+Gender [ADJ] 'bad' – *farr*-Subject+TAM [v] 'be(come) bad'. It is still unclear whether one should assume that inchoative-stative verbs are derived from adjectives, or vice versa, through conversion, or that the word class of property concepts is underspecified, allowing for both adjectival and verbal inflection (Treis 2008: 269).

Kambaata also has five productive derivational processes to create adjectives on the basis of nominal, verbal, ideophonic and adjectival roots (Tab. 3), the first four processes are briefly addressed in this section, the approximative derivation is discussed in detail in the following. All derivational morphology (in bold) is found between the root and the adjectival case/gender inflection; the latter is realised as *-á/-áta*, *-ú/-íta*, *-ú/-úta* in the accusative (= citational) masculine/feminine form, depending on the declension of the resulting derived adjective.

Tab. 3: Productive adjectivising derivations

Derivation	Morphological structure	Example
Proprietive	Noun-***aam***-Case+Gender	e.g. *orc-á* 'mud' > *orc-**aam**-ú* (M)/*-íta* (F) 'muddy'
Agentive	Verb/Noun-***aan***-Number-Case+Gender	e.g. *dimb-* 'be(come) drunk' > *dimb-**aan**-ch-ú(ta)* 'drunkard'
Resultative	Ideophone-***eem***-Case+Gender	e.g. *dákk=y-* 'be hidden' > *dakk-**eem**-á(ta)* 'hidden'
Caritive	Noun-***beel***-Case+Gender	e.g. *wozan-á* 'heart' > *wozan-**beel**-ú(ta)* 'forgetful (lit. heartless)'[6]
Approximative	Adjective/Verb/Noun/etc.-***lab***-Case+Gender	e.g. *mux-á(ta)* 'wet' > *mux-(i)**lab**-á(ta)* 'a bit wet'

Proprietive adjectives attribute to the head noun they modify the property of possessing, being full of, containing or being characterised by N, i.e. the nominal root on which they are based (Treis 2008: 274–277); see, for instance, *buur-ú* 'butter' > *buur-aam-ú/-íta* 'containing butter (e.g. coffee)' and *fool-í* 'soul' > *fool-aam-ú/-íta* 'soul-having (i.e. living)' (9).

---

[6] The heart is the seat of the mind in Kambaata.

Proprietive adjectives can express a time-stable property, e.g. *maq-aam-ú/íta* 'strong (lit. having strength)' < *maq-óo* 'strength', or a temporary property, which enables reference to someone (or something) standing out from a group, e.g. a girl who is *boos-aam-íta* 'water pot-having' (< *boos-ú* 'water pot') is probably accompanied by girls who do not carry a water pot. Proprietive adjectives take the case/gender markers *-ú* (M.ACC)/*-íta* (F.ACC).

(9) *Mann-íi*   *gizz-íi*   *reh-ée=hanníi*
     people-M.DAT   domestic_animal-M.DAT   die-3M.PFV.REL=NMZ3.M.DAT

     *fool-**aam**-íi*   *jabá*
     soul-PROP-M.DAT   bless.INTJ

     'Bless (the coffee) for the people, for the domestic animals, for the dead, for the living (lit. soul-having)!' (AN2016-02-19_001)

The agentive derivation takes verbal (and to a lesser extent nominal) roots as basis for adjectives expressing 'who/which verb-s; who/which is engaged in noun'; see, e.g., *mogga'-aan-ch-ú(ta)* 'who (m/f) steals' < *mogga'-* [verb] 'steal' and *zeem-aan-ch-ú(ta)* 'who (m/f) is engaged in herding' < *zeem-áta* [noun] 'herding'. The referents characterised by these agentive adjectives can be habitual/professional agents just as well as occasional/one-off agents (consult Treis 2011 for other less common meanings of the agentive derivation). The derivational morpheme *-aan* is obligatorily followed by a number marker, either *-ch* SGV or *-n* PLV3, and the case/gender morpheme *-ú* (M.ACC)/*-úta* (F.ACC); see (8).

The resultative derivation takes a simple or reduplicated ideophone as input and conveys that a referent has undergone the process expressed by the ideophone and is therefore now "ideophoned" (Treis 2008: 283–285; forthcoming: Section 3.4.1), e.g. *qonfo'll-eem-á(ta)* 'dented' < *qonfó'll=y-* 'get a dent, be dented' and *luf~luf-eem-á(ta)* 'soft, malleable, fluffy' < *lúf~lúf=y-* 'be(come) soft, malleable, fluffy'.[7] In addition to the 26 examples given in Treis (2008: 283–284; forthcoming: Section. 3.4.1), Alemu (2016) contains 38 resultative derivates as entries or in the monolingual definitions; all *-eem* derivates are deideophonic.

---

[7] Ideophones obligatorily combine with a light verb to inflect. In the examples given here, the enclitic element *=y-* is the light verb 'say'.

The caritive morpheme *-beel* generates denominal adjectives with the meaning 'N-less, not having N'.[8] It either attaches to the nominal root or (with a small number of nouns ending in a long vowel) to the genitive form; it is followed by the case/gender markers *-ú* (M.ACC)/ *-úta* (F.ACC). In Treis (2008: 277), the caritive derivation was too hastily characterised as "seldom used", whereas Alemu's (2016) dictionary shows that it is in fact fairly productively applicable, at least in written Kambaata. See, for instance, *bonx-beel-ú(ta)* 'leafless, without leaves' < *bonx-á* 'leave(s)' and *seer-beel-ú(ta)* 'illegal, without rules' < *seer-á* 'rule, law'. However, most caritive adjectives attested in the written corpus are best considered semantically transparent *ad hoc*-creations and the result of the generalisation of a hitherto only weakly productive derivational schema. Lexicalised caritives that are not *ad hoc* but widely shared across the speech community are, e.g., *wol-beel-ú(ta)* 'countless' < *woll-úta* 'counting, number', *wozan-beel-ú(ta)* 'forgetful (lit. heartless)' < *wozan-á* 'heart', *su'mm-beel-é* 'first week of July (lit. nameless)' < *su'mm-á* 'name', *maq-ee-beel-úta* 'pregnant (lit. powerless)' < *maq-ée* (F.GEN) 'of power', *hagar-beel-ú(ta)* 'indescribable, uncategorisable, ugly (lit. typeless)' < *hagar-á* 'type', *man-beel-(ch)-ú(ta)*[9] 'foreign, without relatives in the area (lit. peopleless)' < *mann-á* 'people'. Many of these established caritives are semantically idiosyncratic.

Finally, Kambaata can also create new adjectives through compounding. Compound adjectives are suprasegmentally and inflectionally indistinguishable from derived adjectives; they differ from the latter only in the fact that compounding combines two lexical roots. One prominent and fairly productive type of compound adjective consists of a numeral plus a nominal root, e.g. *mat-* 'one' + *ill-íta* 'eye' > *mat-ill-á(ta)* 'one-eyed' (see Treis 2017a for numerous examples). A second type combines a nominal root with the verbal root *agud-* 'seem, resemble',[10] e.g. *bun-agud-á(ta)* 'coffee-like, coffee-coloured; (neolog.) brown' (< *bun-á* 'coffee'), *ann-agud-á(ta)* 'resembling one's father' (Alamu 2022: 24) (< *ann-á* 'father') and *am-agud-á(ta)* 'resembling one's mother', an adjective which has

---

[8] In earlier works on Kambaata morphology, the derivation with *-beel* has been called "privative". To avoid confusion with the use of "privative" in the sense of 'fake, non-authentic' in the introduction and other papers in this Special Issue, I have decided to re-label *-beel* a "caritive" morpheme, in accordance with Oskolskaya (2020) and the research project "Grammatical periphery in the languages of the world: A typological study of caritives" (www.caritive.org).

[9] The *-ch* morpheme marks the singulative (SGV).

[10] In Treis & Deginet (2019: 229), adjectives with *-agud* were erroneously considered to be the result of a similative derivational process.

given rise to the proper noun *Am-agud-é*[11] for a cow resembling its mother. Finally, a third compound type links two nominal roots, the second of which is *manka'-á* 'behaviour, manners', as, e.g., in *oos-manka'-á(ta)* 'child-like, young person-like' (< *oos-úta* 'children, young persons'). This compound type will concern us further in Section 5.3. All compound adjectives receive the default case/gender markers of the A1 declension: *-á* (M.ACC)/*-áta* (F.ACC).

5. Approximative derivation

5.1 Morphology

The morphological structure of approximative adjectives is sketched in Fig. 1. The derivational morpheme *-lab* is attached to a simple root or a derived stem.

		Approximative derivation		Case/gender inflection
A – N – V stem		(epenthetic *-i*)	**-lab**	*-á* (M.ACC)
Root	(Derivation)			*-áta* (F.ACC)

Fig. 1: Morphological structure of approximative adjectives

If the morpheme is suffixed to roots or stems ending in a consonant cluster, one either observes epenthesis or cluster simplification to avoid an illicit consonant sequence (10); the choice of either strategy is speaker-dependent, sometimes there is even variation in the pronunciation or spelling of one and the same speaker or author.

(10) *gambal-**lab**-á(ta)* ~ *gamball-i-**lab**-á(ta)* 'blackish' (< *gamball-á(ta)* 'black')
      (Alamu 2022: 7; Alemu 2016: 812)

Approximative adjectives are categorised into declension A1, marked by *-á* (M)/*-áta* (F) in the accusative (= citational) form. The inputs of the approximate derivation are lexemes of different word classes; compared to other adjectivisers (cf. Tab. 3), the approximative derivation has the least selection restrictions. In my database, adjectives, verbs and common nouns constitute about one third of the inputs each (Tab. 4). Inputs from other word classes

---

[11] The case/gender morpheme *-é* is the accusative morpheme of the largest declension of feminine names (Treis 2008: 103).

are only marginally attested; sometimes the word class of the base cannot be unequivocally determined.

Tab. 4: Inputs of the approximative derivation across word classes

Base	Simple	Extended	All
Adjective	22	5	27
Verb	19	4	23
Common noun	27	1	28
Proper noun	1	-	1
Ideophone	1	-	1
Numeral	1	-	1
Undetermined	3	-	3
Total	74	10	**84**

Extended stems on which approximative adjectives are based may be derivates of the agentive (ag) (11), singulative (sgv) (11), middle (mid) (12), reciprocal (mid-pass) (13) and maybe also other, still unattested derivational mechanisms.

(11)   (…) *húj-i-'*   *mereer-aan-ch-i-**láb**-a-ta*
        work-F.NOM-1SG.POSS   middle-AG-SGV-EP-**APRX**-F.PRED-F.COP2

        '(…) my work is fair-to-middling/somehow okay.' (Saint-Exupéry 2018: 21)

(12)   *uucc-aqq-**lab**-á*
        beg-MID-**APRX**-M.ACC

        'almost begging (for one's benefit)'

(13)   *boroo<'>rr-am-**lab**-á*
        insult<MID>-PASS-**APRX**-M.ACC

        'almost insulting (and provoking a fight with) each other'

Like other adjectives, approximative adjectives can themselves be the input of the abstract noun derivation, marked by *-im* (14) or *-oom* (15), which expresses a quality, state or condition.[12]

(14)   *shiish-**lab**-im-áta*
        sour-**APRX**-QUAL1-F.ACC

        'slight sourness, slight acidity' (Alemu 2016: 797)

---

[12] The morphemes *-im* and *-oom* are rivaling suffixes for the derivation of quality nouns. It is not yet clear what determines the use of either morpheme in this function. The suffix *-oom* has a second function of deriving status nouns, e.g. *meent-oom-áta* 'womanhood' < *meent-ú* 'women' (Treis 2008: 154–156, 171–174).

(15)   *mux-**lab**-oom-áta*
         wet-**APRX**-QUAL2-F.ACC

         'dampness, slight wetness' (Alemu 2016: 821)

5.2 Morphosyntax

Just like other simple and derived adjectives (Section 3), approximative adjectives are used in four syntactic functions. As adnominal modifiers in the NP, they agree in case and gender with the head noun, e.g. with a masculine nominative subject ('leaf') in (16) and, in its oblique form, with a feminine locative adjunct ('on feather(s)') in (17); recall also (1) and (3).

(16)   *bull-i-**láb**-u*   *bónx-u*   *yoo-sí*
         grey-EP-**APRX**-M.NOM   leaf-M.NOM   COP1.3-3M.OBJ.REL

         biix-am-ú   *sarb-umb-ú*   *haqq-á*
         break-PASS-M.ACC   do_fast-3M.NEG5-M.ACC   tree-M.ACC

         (Definition of a plant species:) 'a tree that has greyish leaves and does not break easily' (Alemu 2016: 759)

(17)   *xabar-**láb**-ata*   *ball-éen*   *wójj-u*
         ash-**APRX**-F.OBL   feather-F.LOC   white-M.ACC

         xéf~xéf=y-ee-sé-e (…)
         RED~be_speckled.IDEO=say-3M.PFV-3F.OBJ.REL-NMZ1.M.ACC

         (Excerpt of the definition of "guinea fowl":) 'white speckles being on ash-like feathers (…)' (Alemu 2016: 1157)

Approximate adjectives can become heads of NPs without further measures (such as nominalisation) being taken; see (18) and further on (39).

(18)   *ang-áta*   *aa'll-eemmá*   *zakk-óon*   *moosh-sh-íi*
         hand-F.ACC   wash.MID-3HON.PFV.REL   after-M.LOC   dry-CAUS1-M.DAT

         dagla'll-eennó   *oddish-**lab**-á*
         use-3HON.IPFV.REL   clothes-**APRX**-M.ACC

         (Definition of "towel":) 'a clothes-like (thing) which one uses to dry (one's) hands after one has washed them' (Alemu 2016: 319)

In their third syntactic function, as heads of non-verbal predicates, they agree in gender with the overt or retrievable subject and combine with the ascriptive-identificational

copula (COP2);[13] the masculine copula *-(h)a(a)* in (19) is determined by the masculine sub-ject 'father', the feminine form *-ta(a)* in (11) by the feminine subject 'work'.

(19) *Ánn-u*   *qeraa'rr-i-**láb**-a-a,*
      father-M.NOM   tall-EP-**APRX**-M.PRED-M.COP2

      béet-u-s-ma   *gabbán-ch-u-a*   *bagáan*
      son-M.NOM-3M.POSS-CFOC   short-SGV-M.PRED-M.COP2   CNTR

      (A says: 'All family members are short.' – B corrects him/her:) 'The father is a bit tall, his *son* is short, though.' (Elicited, 3/2018)

Finally, approximative adjectives can be used as adverbial modifiers. As such, they are in-variantly marked for the masculine accusative case. The accusative does not only mark direct objects but also certain adverbial constituents, see, e.g., the simple adjective *qahúnka* '(for) a little (while)' in adverbial function in (24). In (20), the approximative adjective is an adverbial modifier to a subordinate verb (converb). See also (2).

(20) *Sikk-i-**lab**-á*   *maram-án*   *márr-ee-haa,*
      drag_oneself-EP-**APRX**-M.ACC   walk-3M.IPFV.CVB   go-3M.PRF.REL-M.COP2

      zuuq-qáa   *man-ch-ó=b-a*
      massage-3F.IPFV.REL   person-SGV-F.GEN=PLC-M.ACC

      'He almost dragged himself to the masseuse.' (Elicited, 3/2018)

The word class of the base has no influence on the syntactic function that the approxima-tive adjective can adopt. Irrespective of whether the base is an adjective, verb or noun, the derived form can be used in all syntactic functions permitted to adjectives. In (21), we see the use of a deverbal approximative as a modifier in a noun phrase. This is to be compared with (16) and (17), which show deadjectival and denominal approximatives in modifying function in the NP.

(21) *geeq-**láb**-ata*   *ill-éen*   *xuud-ú*
      stare-APRX-F.OBL   eye-F.ICP   look-M.ACC

      'to look in a kind of staring way (lit. look with almost staring eyes)' (Alemu 2016: 535)

---

[13] Kambaata has four copulas whose intricate distribution is explained in Treis (2008: chap. 10).

After having shown in this and the preceding section that the outputs of the approximative derivation are undoubtedly adjectival in nature, I turn to the semantics of *-lab* in the next two sections.

5.3 Meaning

The analysis of the meaning of the approximative derivation takes a native speaker's perspective as a preliminary starting point. In Alemu's dictionary (2016: 635), we find an entry *-laba*, labelled *B*(*otokonnuta*) '(neolog.) affix', defined as in (22).

(22)	*su'm[m]-caakk-is-aan-ch-óon*			*botokonn-ámm*
	noun-become_light-CAUS1-AG-SGV-M.LOC			attach-PASS.3M.PFV.CVB

	kotim-á	*bikk-á*	*ih-ú*	*kul-is-aan-ch-ú*
	less-M.ACC	amount-M.ACC	become-M.ACC	tell-CAUS1-AG-SGV-M.ACC

(Definition of *-laba*:) '(something) that signals a lesser amount (/degree) (when) attached to an adjective (lit. noun-enlightener)' (Alemu 2016: 635)

The monolingual definition points out two important aspects: the approximative derivation takes adjectives as its base and expresses a reduced (attenuated) degree of a property. Both statements need to be qualified. Firstly, as has been shown in Section 5.1, adjectival roots are not the only possible inputs. In the following, I will show, furthermore, that attenuation is only one of several related meanings of the approximative derivation. The meanings of the approximative are dependent on the word class of the base but seem close enough to be subsumable under an overarching meaning of approximation, understood as deviation from a prototypical realisation.

5.3.1 Attenuated property

The first two attestations of approximative forms that occurred in a recording of natural speech are utterances with the colour adjective 'yellowish' (23).

(23)	*Baar-**lab**-áta*	*ík-k*	*fad-dóo'u,*
	yellow-**APRX**-F.ACC	become-3F.PFV.CVB	go_away-3F.PFV

	hikkánne-n
	P_DEM2.M.OBL-EMP

'She (= the chameleon) became yellowish, just there.' (TD2016-02-11_001)

Colour terms are among the most frequent bases of *-lab* in written texts, with 'reddish' at the top of the list in Table 5.

Tab. 5: Attestations of attenuated colour adjectives in written texts

Colour adjective	Translation	Source
*biiishsh-(i)-**lab**-á(ta)*	'reddish'	(Saint-Exupéry 2018: 19; Alemu 2016: xxvi, 29, 128, 216, 229, 261 307, 225, 842, 844, 1010, 1045)
*gamball-(i)-**lab**-á(ta)*	'blackish'	(Alemu 2016: 464, 812, 868; Alamu 2022: 7)
*bull-(i)-**lab**-á(ta)*	'greyish'	(Alemu 2016: 161, 759, 778)
*wojj-(i)-**lab**-á(ta)*	'whitish'	(Alemu 2016: 160, 1061)
*bun-agud-**lab**-á(ta)*	'brownish'	(Alemu 2016: 163)

Although colours are the most common adjectival inputs, also other property lexemes undergo the approximative derivation, e.g. 'wet' in (24), 'fat' in (25) and 'poor' in (26).

(24)   A: *Oddíshsh-at*   *tadá*   *mool-táa'u,*   *aag-ís-i!*
                clothes-F.NOM   by_now   become_dry-3F.IPFV   enter-CAUS1-2SG.IMP

           B: *Mux-i-**láb**-a-a<n>ta,*   *qah-ú<n>ka*
                wet-EP-**APRX**-F.PRED-F.COP2<EMP>   small-M.ACC<EMP>

                hád-a-n   *egér-tun*
                outside-F.OBL-EMP   wait-3F.JUS

A: 'The clothes should be dry by now, bring (them) in!'
B: (goes to check): 'They are (still) a bit wet, leave (them) outside for a little (while).' (Elicited, 3/2018)

(25)   *Āā,*   *gaan-**láb**-a-ta,*   *mereer-áan-ch-u-ta*
        yes.INTJ   fat-**APRX**-F.PRED-F.COP2   middle-AG-SGV-F.PRED-F.COP2

(Context: A asks whether a person is corpulent. – B replies:) 'Yes, she is somewhat fat, middle(-sized).' (Elicited, 6/2022)

(26)   *Kám,*   *bux-ichch-i-**láb**-a-a<n>ka,*
        INTJ[objection]   poor-SGV-EP-**APRX**-M.PRED-M.COP2<EMP>

        áff-ee-si   *bors-á*   *xuud-dáanti-bay?!*
        seize-3M.PRF-DEF.REL   bag-M.ACC   see-2SG.IPFV-BAY

(Context: A says: 'He wears a beautiful coat, he must be rich.' – B replies:) 'No! He is fairly poor, you see the bag that he carries, don't you?' (Elicited, 3/2018)

On a scale from the lowest to the highest degree of a property, deadjectival approximatives cover a vague domain between the two poles. While the *-lab* derivate in (24) expresses a low degree of wetness, i.e. 'a bit wet, almost dry' (see also the recommendation to wait 'a little' for it to dry), the rephrase that follows the *-lab* derivate in (25) shows that the speaker considers the degree of fatness to be in the middle of two body size extremes. Many deadjectival approximatives express a degree that is slightly reduced when compared to the norm: the attenuated colour in (23) is not a full, prototypical yellow, and the person in (26) is not poor but close to being poor. Deadjectival approximatives have been paraphrased, as in (27), with an adverbial *qah-únka* 'a bit'.

(27) *Ánn-u*   *oos-óochchi-s*   *qah-ú<n>ka*
      father-M.NOM   children-F.ABL-3M.POSS   small-M.ACC<EMP>

      qeráa'rr-u-a
      tall-M.PRED-M.COP2

      Solicited periphrasis of (19): 'The father is a bit taller than his children (= who are all short).' (Elicited, 6/2022)

Many deadjectival approximatives confirm Alemu's (2016: 635) intuition, as cited in (22), in that they express an attenuated degree of a property. As the position of approximative-marked properties on a scale between non-realisation and prototypical realisation is vague, they are translated, depending on the context, as 'a bit [PROPERTY], [PROPERTY]-ish, fairly [PROPERTY], almost [PROPERTY]'.

5.3.2 Similarity and near-identity in manner

In my database, 23 different deverbal adjective types are attested, of which 16 are found in Alemu (2016). A selection of simple and extended verbal stems is presented in (28)–(30). Note that a sequence of a middle and a passive morpheme as in (30) marks the reciprocal (Treis in press: Section 4.4).

(28) *maqees-**lab**-á*   'in a kind of forceful way'   < *maqees-* 'force (tr.)'
      *gis-(i)-**lab**-á*   'in a kind of sleeping way'   < *gis-* 'sleep'
      *uucc-**lab**-á*   'in a kind of begging way'   < *uucc-* 'beg'

(29)   *uucc-aqq-**lab**-á*

     beg-MID-**APRX**-M.ACC

     'almost begging (for one's benefit)'

(30)   *boroo<'>rr-am-**lab**-á*

     insult<MID>-PASS-**APRX**-M.ACC

     'almost insulting (and provoking a fight with) each other'

The use of deverbal approximatives in context is illustrated in (31)–(32). The approximative derivation has little influence on the verbal argument structure. Even after having undergone adjectivisation, the verbal base *uucc-aqq-* 'beg for one's benefit' in (31) can still govern an accusative direct object, the tea house boy.[14]

(31)   *Ciil-íi*   *daabb-úta*   *háshsh*   *bors-áan-ta-'*

     infant-M.DAT   bread-F.ACC   want.1SG.PFV.CVB   bag-M.LOC-LNK-1SG.POSS

     gizz-á   *xuujj-oommí=da*   *bácc-o.*

     money-M.ACC   see-1SG.PFV.REL=COND   be_absent-3M.PFV

     Shay-í   *min-í*   *adab-áa*   *uucc-aqq-(i)-**lab**-á*

     tea-M.GEN   house-M.GEN   boy-M.ACC   beg-MID-EP-**APRX**-M.ACC

     ass-eemmi=dá-a   *kámm-o-'e*

     do-1SG.PFV.REL=COND-ADD   reject-3M.PFV-1SG.OBJ

     (Context: Person reports about an incident at the bakery) 'I wanted to buy bread for the child, but when I looked for money in my bag, there was none. Even though I kind of/almost begged the tea house boy (i.e. the seller), he rejected (to give the bread) to me.' (Elicited, 3/2018)

(32)   *M-á*   *y-ee=dá-a*   *dag-im-bá'a.*

     what-M.ACC   say-3M.PFV=COND-ADD   know-3M.NIPFV-NEG1

     *Gis-i-**lab**-á*   *ih-áne-et*

     sleep-EP-**APRX**-M.ACC   become-1SG.IPFV.CVB-COP3

     maccoocc-eemmí-ihu

     hear-1SG.PRF.REL-NMZ1.M.NOM

     (Context: A: 'What did he say?' – B:) 'I don't know whether he said anything. I heard (something but) I was kind of sleeping.' (Elicited, 3/2018)

---

[14] Deverbal agentive adjectives also continue to govern accusative objects (Treis 2012: 242–243). It is unknown whether deverbal approximative (like agentive) adjectives can also govern adverbial constituents.

As seen in (31)–(32), deverbal -*lab* adjectives express an approximate, often near-identical manner of doing something, a 'kind of VERB-ing' or an 'almost (but not quite) VERB-ing manner'. Whereas deadjectival forms (Section 5.3.1) convey a vague, more or less reduced degree of a property, the relationship between the event expressed by the verbal base *vs.* the event expressed by the deverbal approximative is a relationship of close similarity or near-identity. Deverbal approximatives are approximating in the literal sense, in that they come close to the prototypical realisation of an event. The near realisation of an event expressed by a deverbal approximative is especially evident with achievement verbs such as *ub-* 'fall', as in (33) about a near-accident situation.

(33) *Órc-u*   *shúrr=a'-í-yan-s*
      mud-M.NOM   slip.IDEO=do-3M.PFV.CVB-DS-3M.OBJ

      *ub-**lab**-á*   *íkk*   *fanqáll-o*
      fall-**APRX**-M.ACC   become.3M.PFV.CVB   return-3M.PFV

      'The mud made him slip, he almost fell down (but) caught himself.' (Elicited, 6/2022)

The appropriateness of the translation 'almost VERB-ing' is also supported by the periphrasis of (34) in (35). Here a deverbal approximative is paraphrased with the verb *hoog-* 'miss, not do, fail to realise (here: the mistake)'.

(34) *Luus-**lab**-á*   *íkk*   *fanqall-óomm*
      make_a_mistake-**APRX**-M.ACC   become.1SG.PFV.CVB   return-1SG.PFV

      'I almost made a mistake (but then) changed course (lit. returned).' (Elicited, 3/2018)

(35) *Lúushsh*   *hóogg*   *dagg-óomm*
      make_a_mistake.1SG.PFV.CVB   miss.1SG.PFV.CVB   come_to_know-1SG.PFV

      Periphrasis of (34): 'I almost made a mistake (but then) became aware (of it) (lit. I made a mistake, missed, come to know).' (Elicited, 3/2018)

5.3.3 Similarity in quality or character

Denominal approximative adjectives characterise something as 'similar to/almost like noun in quality or character, of the same type'. My database contains 28 different denominal adjective types, of which some are attested several times; see (36)–(37). Denominal approximatives create *ad hoc* categories (in the sense of Mauri 2017, to which the reader is

referred for a typological survey of the linguistic means to construct *ad hoc* categories). *Ad hoc* categorization is an exemplar-driven, context-dependent abstraction process. Based on one salient exemplar, e.g. the weteechchu-tree in (36), which can be assumed to be known to all Kambaata speakers, a superordinate category is built for which no established term exists in Kambaata, i.e. the class of objects that are like the weteechchu-tree. This novel superordinate category includes both the explicit exemplar and other members that share a specific property with it. The property has to be inferred by the interlocutor from the linguistic or extra-linguistic context; in the case of weteechchu-like objects, the culturally salient property is having slow-burning fragrant wood.

(36)   *wet-eechch-**lab**-á*   *haqq-á*
         tree_sp-SGV-**APRX**-M.ACC   tree-M.ACC

         'a *weteechchu*-like tree' (Alemu 2016: 147, used for the definition of the lesser-known tree species *boolima*)

In (37), an *ad hoc* category of objects made from fabric is established, based on clothes as the salient exemplar with this particular property. The approximative derivation is amply exploited in the monolingual dictionary definitions in Alemu (2016), where the *ad hoc* categories based on a salient member are recruited to characterise other, less-known members of the category, e.g. the *boolima*-tree, a tree that also has fragrant wood (36), and a towel, a modern-day object, which is made of the same material as clothes (37).

(37)   *oddishsh-**lab**-á*
         clothing-**APRX**-M.ACC

         'a clothing-like, garment-like (thing)' (Alemu 2016: 319, used in the definition of a towel)

Denominal approximatives are used as case/gender-agreeing modifiers in the NP in (36) and (38) and as NP heads in (37) and (39). They can equally well be used as predicates or adverbial modifiers.

(38) *Oot-á* *al-íichch* *marf-**lab**-á* *ut-ichch-ú*
      hedge-F.GEN top-M.ABL needle-**APRX**-M.ACC thorn-SGV-M.ACC

      áff *wáall-o*
      seize.3M.PFV.CVB come-3M.PFV

      'He brought a needle-like thorn from the hedge (e.g. to take out a splinter).' (Elicited, 3/2018)

(39) *Ann-i-sí* *urr-óon* *min-**lab**-á*
      father-M.GEN-3M.POSS front_yard-F.LOC house-**APRX**-M.ACC

      uurr-íshsh-ee'u
      stand-CAUS1-3M.PRF

      'He has built a kind of a house/a house-like thing (= not a proper house in the speaker's mind) in his father's front yard.' (Elicited, 3/2018)

Denominal approximatives are paraphrasable by similative constructions, as two consulted native speakers pointed out (40). Kambaata has an enclitic =g- 'manner; like' (see Treis 2017b; Treis 2018 for detailed descriptions), which marks, among other, the standard of comparison in similative and equative constructions.

(40) *wet-eechch-í=**g**-a* *yóo* *haqq-á*
      tree_sp-SGV-M.GEN=SIM-M.ACC/OBL COP1.REL tree-M.ACC

      Periphrasis of (36): 'a tree which is like the *weteechchu*-tree' (Elicited, 6/2022)

Denominal approximatives tend to express a similarity in nature, character or function and less so in physical appearance. This observation is captured, for instance, in contrasting pairs such as the (neo-)compound adjective *bun-agud-á(ta)* 'coffee-coloured' (< *bun-* 'coffee' + *agud-* 'resemble') and the approximative adjective *bun-(i)-lab-á* 'coffee-like', of which the first form – according to native speaker intuitions – is preferred to characterise the colour of an object, while the second form is preferred to characterise objects that have the same function, significance or way of consumption.

5.3.4 Negative connotations and semantic idiosyncrasies

Even though approximatives can have a negative connotation, as seen in (39) where the speaker belittles somebody's house through the use of *min-lab-á* 'kind of house, house-like (thing)', they most often seem to have a neutral connotation, as the numerous other examples in this contribution show.

The approximative derivation is semantically regular and the meaning of derivational outputs predictable. This stands in contrast to other adjectivising derivations that display a certain degree of semantic idiosyncrasy – see, for instance, the caritive adjectives and their idiosyncratic translations mentioned in Section 4. So far only one approximative form is attested in an idiomatic expression, see the denumeral approximative in (41).

(41)   *Xah-á-s*   *xább=át-t*
        thing-M.ACC-DEF   do_well.IDEO=do-3F.PFV.CVB

        makk-is-sim-bá'a,   *mát-u*
        be_convenient-CAUS1-3F.NIPFV-NEG   one-M.NOM

        *tordum-**láb**-a-a=rr-a*
        ten-**APRX**-M.OBL-COP2=NMZ4-M.PRED

        (Possible context: conversation about a disorganised event) 'They have not arranged the things (e.g. the talks) well, everything is a mess (lit. one is ten-like ones).'
        (Elicited, 3/2018)

One other deverbal approximative is known to have a (slightly) unpredictable figurative meaning (42), which possibly motivated Alemu to integrate it into the dictionary (2016: 1002).

(42)   *Torr-i-**lab**-á*   *hírr*   *waall-óomm*
        throw-EP-**APRX**-M.ACC   sell.1SG.PFV.CVB   come-1SG.PFV

        'I sold it very cheap (lit. almost like throwing) and came back (home).'
        (Elicited, 3/2018)

5.4 Competing strategies

In the preceding sections, periphrases of approximative forms have helped shed light on certain aspects of the meaning of the *-lab* derivation. In this section, I briefly address near synonyms of approximatives as they are attested in Alemu (2016). In the Kambaata dictionary, compound adjectives of which the second element is the nominal root of *manka'-á* (M) 'humility, behavior, manner(s), norm(s)' sometimes occur in the same contexts or as periphrases of denominal approximatives (Section 5.3.3). Compare the descriptions of two bird species with ash-coloured feathers in (43) and (44), one of which uses a compound adjective with *manka'-*, the other a derived adjective with *-lab*.

(43)   *xabar-**manká'**-u*   *hagár-u*   *yoo-sé*
       ash-**manner**-M.NOM   colour-M.NOM   COP1.3-F.OBJ.REL

          cii'-á   *sheef-á*
          bird-F.GEN   species-M.ACC

          (Definition of *handarchúta* 'pigeon':) 'a bird species of ash-like colour'
          (Alemu 2016: 439)

(44)   *xabar-**láb**-at*   *báall-it*   *yoo-sé*   *zigraa'-í*
       ash-**APRX**-F.NOM   feather-F.NOM   COP1.3-3F.OBJ.REL   guinea_fowl-M.GEN

          sheef-á   *ík-k*   *gabb-áta*   *ciichch-úta*
          species-M.ACC   become-3F.PFV.CVB   mid_sized-F.ACC   bird.SGV-F.ACC

          (Definition of *heedoorchúta* 'partridge':) 'a mid-sized bird of the guinea-fowl family
          with ash-like feathers' (Alemu 2016: 340)

We also find two dictionary entries of attenuated colours. In the first entry, a *manka'*-compound is defined by a deadjectival approximative (45); in the other case, a deadjectival approximative is defined by a *manka'-* compound (Alemu 2016: 161).

(45)   *baar-**manka'**-á*   [Definition:]   *feeg-á*   *baar-á*
       yellow-**manner**-M.ACC   bright-M.ACC   yellow-M.ACC

          ih-umb-ú;   *baar-**lab**-á*
          become-3M.NEG5-M.ACC   yellow-**APRX**-M.ACC

          'yellowish' [Definition:] 'not being bright yellow, yellowish' (Alemu 2016: 87)

Alemu (2016) contains altogether 40 *manka'-* compounds, none of which have a verb as first element.[15] No *manka'-* compounds are attested in my natural fieldwork data. The vast majority of the similative and attenuative *manka'-* compounds in the dictionary are unheard of by my language assistant (Deginet Wotango Doyiso, personal communication); he provided, however, the examples in (46)–(47) in order to illustrate the use of two compounds that he believed to be shared by the speaker community. It seems possible that *manka'* compounds only compete with approximatives in certain regiolects or sociolects of Kambaata.[16]

---

[15] Two compounds are also attested in schoolbooks: *oos-manka'-á(ta)* 'child-like, young person-like', *meneraal-manka'-á(ta)* 'mineral-like' (Kambaata Education Bureau 1989: 6.121; 8.100, 102, 104).

[16] Interestingly, the noun *manka'a* 'behaviour, norm' is found in the glossary of Tadesse's (2015: 339) Hadiyya grammar – which could mean that *manka'á* is a Hadiyya loanword used in a certain geographical area or by a certain Kambaata speaker group. Until some decades ago, Kambaata was under strong cultural and political influence from Hadiyya, which is reflected, among others, in Hadiyya personal names in today's

(46)   *Oos-**manká'**-a-a<n>ka,*
       children-**manner**-M.PRED-M.COP2<EMP>

      batá'   *y-ée=da*   *lamoodúm-a-at*
      become_many.3M.PFV.CVB   say-3M.PFV.REL=COND   twenty-M.OBL-COP3

      (Possible context: How old could he be? – Answer:) 'He is (still) child-like/young person-like, he is at most 20 (years old).' (Elicited, 11/2019)

(47)   *Beenaar-í*   *wud-íin*   *yóo*   *mánn-u*
      B.-M.GEN   side-M.ICP   COP1.3.REL   people-M.NOM

      *Hadiy-**manká'**-a-a*
      Hadiyya-**manner**-M.PRED-M.COP2

      'The (Kambaata) people in the Beenaara area are Hadiyya-like/half-Hadiyya.' (Elicited, 11/2019)

Data in Alemu (2016) also illustrates a strategy that potentially competes with verb-based approximatives. Verbs that combine with an undetermined ending *-xaachch-á* are found to express 'almost verb-ing' (see Section 5.3.2), e.g. *ub-xaachch-á* 'almost fall' (Alemu 2016: 335). However, this particular word formation process and the examples given could not (yet) be confirmed – which leaves the question of competing morphological strategies of the *-lab* derivation open for future research.

6. Diachrony

The Kambaata approximative morpheme *-lab* is most likely linked to a verb root *lab-* that we find in related Highland East Cushitic languages; see *lab-* 'resemble, seem' in Gedeo [drs], Sidaama [sid], Hadiyya [hdy] and Libido [liq] (Hudson 1989: 123; Crass n.d.). The verb root is not (or: no longer) known in Kambaata and, to the best of my knowledge, also unattested in Alaaba and K'abeena (Kambaata's closest relatives; shared ISO-code 639-3: [alw]).[17] The meaning 'resemble, seem' is instead expressed by *agud-* in Kambaata (48), a

---

oldest generation and in loanwords in certain semantic domains. Hadiyya loanwords are especially common in the Kambaata spoken along the borders to Hadiyya.

[17] Kambaata has a verb *lab-* 'lure; woo, court (a woman)', which is not semantically related but merely homophonous to Highland East Cushitic *lab-* 'resemble, seem' (Deginet Wotango Doyiso pers. comm.; Alemu 2016: 635).

verb which is not attested in Highland East Cushitic languages outside the Kambaata-Alaaba-K'abeena group.[18]

(48) *Bahír-u-s*   *cíil-u*   *ann-íichchi-s*   *abb-á*
      elder-M.NOM-DEF   infant-M.NOM   father-M.ABL-3M.POSS   more-M.ACC

      am-á-s   **agud**-*áno-a*
      mother-F.ACC-3M.POSS   **resemble-**3M.IPFV.REL-M.COP2

      'The eldest son resembles his mother more than his father.' (Elicited, 9/2021)

The Kambaata approximative derivation in the synchronic Stage 2 probably has its origin in a compounding process that combined two lexical roots, the second of which was the verb **lab-* 'resemble' (49), cf. Stage 1.

(49)   Possible historical development:
          Stage 1   →   Stage 2
          Compound adjective: Derived adjective
          Root 1 + Root 2 (**lab-*) + Inflection   Root 1 + APRX-Derivation + Inflection

A parallel compounding process is attested with the verb root *agud-* 'resemble' in synchronic Kambaata – see the adjectival noun-verb compounds like the neologism *bonx-agud-á(ta)* '(neolog.) green' (< *bonx-á* 'leaves' + *agud-* 'resemble') and similar examples mentioned at the end of Section 4. Like all compound adjectives, the *-lab* derived approximative adjectives belong to declension A1, which is another small piece of evidence that the latter might have developed out of compounds.[19]

When at one point in time the verb stem **lab-* 'resemble' was no longer used on its own, became obsolete and/or was ousted by *agud-*, the element *-lab* in compounds turned from a verbal root into a derivational affix. The change in morphological status and the loss of the link to an existing lexeme was possibly accompanied by semantic bleaching and abstraction as well as an extension of use, which allowed the *-lab* derivation to take lexemes of different word classes as its input. The details of this process remain, however, speculative and undocumented.

---

[18] There is, possibly, a historical link between Kambaata *agud-* 'resemble, seem' and the Sidaama similative morpheme *-gede*, that we find, for instance, in examples in Kawachi (2007).
[19] Note, however, that A1 is also by far the biggest adjectival declension.

The compounding process with *agud-* 'resemble' is still transparent to modern Kambaata speakers. Interestingly, coinages of colour terms that are *agud-* compounds can serve as the input of the approximative derivation, see (50), which shows that speakers do not consider *agud-* and *-lab* to be functionally equivalent.

(50)   *bun-agud-**lab**-á*
        coffee-seem-**APRX**-M.ACC
        'brownish (lit. coffee-seeming-ish)' (Alemu 2016: 163)

The contributions assembled in this special issue discuss different source domains for approximative morphology (the reader is invited to consult the introduction to this issue for an overview and references to the individual chapters). I have shown in this section that the diachronic source of the Kambaata approximative morpheme *-lab* goes back to a verb expressing resemblance. Regarding its origin, the Kambaata approximative is thus comparable to other approximative markers that have developed from items expressing resemblance or similarity, such as the English *-like* suffix and the German *-ähnlich* and *-gleich* suffixes (Hüning & Schlücker 2023).

7. Summary

The Kambaata approximative derivation is a productive, albeit less frequently attested adjectivising derivational mechanism if compared, for instance, to the very common proprietive and agentive derivations. However, Alemu (2016) makes extensive use of the approximative derivation in the monolingual definitions of the entries in his dictionary, and native speakers are able to spontaneously come up with new examples. The fair degree of productivity of the approximative derivation helped base the analysis in this paper on a total of 84 different derivates.

The approximative derivation has few (if any) selection restrictions for the base to which it attaches. The outputs of the derivational process are clearly adjectival in nature, as shown by their agreement behaviour. Approximative adjectives belong to the largest adjectival declension. Syntactically, approximatives, just like basic adjectives, can be used as modifiers in the NP, as heads of the NP, in predicative and in adverbial function, irrespective of the word class of their base.

The meaning of the approximative derivation is, to a certain extent, dependent on the word class of its input. With adjectival bases, approximative adjectives express a vague, more or less attenuated degree of a property, deverbal approximatives express a similar, often near-identical realisation of an event, and denominal approximates express a similarity in nature, character or function and create *ad hoc* categories. The meanings are subsumable under an overarching meaning of approximation, understood as deviation from a prototypical realisation. It is unclear whether Kambaata has word formation processes that compete with the approximative derivation – the allegedly synonymous formations attested in Alemu (2016) could not (yet) be confirmed.

The Kambaata approximative morpheme must have emerged – in the recent history of the language – from an independent lexeme, **lab-* 'resemble', which is no longer in use in the language but still found in related Highland East Cushitic language. From a Cushitic perspective, the Kambaata approximative derivation is unique of its kind. No related language has so far been reported to have a morphological process to express approximation.

Abbreviations

1	1st person
2	2nd person
3	3rd person
A_	adjectival
ABL	ablative
ACC	accusative
ADD	additive ('also', 'and')
AG	agentive
APRX	approximative
BAY	negative rhetorical question
CAUS1	simple causative
CFOC	contrastive focus
CNTR	contrastive
COND	conditional
COP1	existential-locative copula *yoo-*
COP2	ascriptive-identifiational -(*h*)*a*(*a*)-/-*ta*(*a*)-copula
COP3	ascriptive-identifiational V*t*-copula
CVB	converb
DAT	dative
DEF	definite

DEM1   proximal demonstrative
DEM2   medial demonstrative
DEM4   distal demonstrative
DS   different subject
EMP   emphasis
EP   epenthesis
F   feminine
GEN   genitive
HON   honorific, impersonal
ICP   instrumental-comitative-perlative
IDEO   ideophone
IMP   imperative
INTJ   interjection
IPFV   imperfective
JUS   jussive
LNK   linker
LOC   locative
M   masculine
MID   middle
NEG1   standard negator
NEG3   jussive negator
NEG5   relative negator
NIPFV   non-imperfective
NMZ1   nominaliser -V
NMZ3   nominaliser =*hann*
NMZ4   nominaliser =*r*
NOM   nominative
OBJ   object
OBL   oblique
P_   pronominal
PASS   passive
PFV   perfective
PL   plural
PLC   place nominaliser =*b*
PLV3   plurative -*n*
POSS   possessive
PRED   predicative
PRF   perfect
PROP   proprietive
QUAL   quality noun derivation
RED   reduplication
REL   relative
SG   singular
SGV   singulative

References

Alamu Banta & Alamaayyo G/Tsiyoon. 2017. *Hambarrichcho yaanata: Kambaatissa-Amaarsa hayyo'ooma yannaakkata* [Spices of Mount Hambarrichcho: Kambaata-Amharic proverbs]. Addis Ababa: Addis Ababa University.

Alamu Banta Ataara. 2022. *Daambiliiqa: Kambaatissa finka* [Intellectual advancement: Kambaata metaphors]. Duuraame: no publisher.

Alemu Banta Atara [Alamu Banta Ataara]. 2016. *Kookaata: Kambaatissa-Amaarsa-Ingiliizissa laaga doonnuta* [Kambaata-Amharic-English dictionary]. Addis Ababa: Berhanena Selam Printing.

Banti, Giorgio. 1988. "Adjectives" in East Cushitic. In Marianne Bechhaus-Gerst & Fritz Serzisko (eds.), *Cushitic – Omotic. Papers from the International Symposium on Cushitic and Omotic Languages, Cologne, January 6-9, 1986*, 203–259. Hamburg: Buske.

Central Statistical Agency. 2007. Ethiopia: Population and housing census of 2007. http://catalog.ihsn.org/index.php/catalog/3583 (accessed 4 January 2020).

Crass, Joachim. no date. [Collection of transcribed, glossed and translated Libido texts].

Hudson, Grover. 1989. *Highland East Cushitic dictionary*. Hamburg: Buske.

Kambaata Education Bureau. 1989. *Kambaatissata: Rosaanchi maxaafa* [Kambaata language: School book]. 8 vols. Hawassa: Southern Nations, Nationalities, and Peoples Regional State.

Kawachi, Kazuhiro. 2007. *A grammar of Sidaama (Sidamo), a Cushitic language of Ethiopia*. Buffalo: State University of New York at Buffalo (PhD dissertation).

Mauri, Caterina. 2017. Building and interpreting ad hoc categories: A linguistic analysis. In Joanna Blochowiak, Cristina Grisot, Stephanie Durrleman & Christopher Laenzlinger (eds.), *Formal Models in the Study of Language*, 297–326. Cham: Springer. DOI: 10.1007/978-3-319-48832-5_16.

Mous, Maarten. 2012. Cushitic. In Zygmunt Frajzyngier & Erin Shay (eds.), *Afroasiatic languages*, 342–422. Cambridge: Cambridge University Press.

Oskolskaya, Sofia A., Natalia M. Zaika, Sergei B. Klimenko & Maksim L. Fedotov. 2020. Opredelenie karitiva kak sravnitel'nogo ponjatija [Defining caritive as a comparative concept]. *Voprosy Jazykoznanija* 2020(3). 7–21. DOI: 10.31857/S0373658X0009370-6.

Saint-Exupéry, Antoine de. 2018. *Qakkichchu Laaha* [The Little Prince]. (Trans.) Deginet Wotango Doyiso & Yvonne Treis. Neckarsteinach: Tintenfaß.

Shay, Erin. 2014. Afroasiatic. In Rochelle Lieber & Pavol Stekauer (eds.), *The Oxford Handbook of Derivational Morphology*, 573–590. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199641642.013.0032.

Tadesse Sibamo Garkebo. 2015. *Documentation and Description of Hadiyya (a Highland East Cushitic language of Ethiopia)*. PhD dissertation. Addis Ababa: University of Addis Ababa.

Treis, Yvonne. 2008. *A grammar of Kambaata. Part 1: Phonology, Morphology, and Non-verbal Predication*. Cologne: Köppe.

Treis, Yvonne. 2011. Polysemous agent nominals in Kambaata (Cushitic). *Sprachtypologie und Universalienforschung* 64(4). 369–381. DOI: 10.1524/stuf.2011.0026.

Treis, Yvonne. 2012. Categorial hybrids in Kambaata. *Journal of African Languages and Linguistics* 33 (2). 215–254. DOI: 10.1515/jall-2012-0009.

Treis, Yvonne. 2014. Number in Kambaata. In Anne Storch & Gerrit J. Dimmendaal (eds.), *Number: Constructions and Semantics. Case Studies from Africa, Amazonia, India & Oceania*, 111–133. Amsterdam: Benjamins. DOI: 10.1075/slcs.151.05tre.

Treis, Yvonne. 2017a. "They are only two, like the teats of a donkey": Kambataa denumerals revisited. In Raija Kramer & Roland Kießling (eds.), *Mechthildian Approaches to Afrikanistik: Advances in Language Based Research on Africa: Festschrift für Mechthild Reh*, 339–366. Cologne: Köppe.

Treis, Yvonne. 2017b. Similative morphemes as purpose clause markers in Ethiopia and beyond. In Yvonne Treis & Martine Vanhove (eds.), *Similative and Equative Constructions: A Cross-linguistic Perspective*, 91–142. Amsterdam: Benjamins. DOI: 10.1075/tsl.117.05tre.

Treis, Yvonne. 2018. Comparison in Kambaata: Superiority, equality and similarity. *Linguistic Discovery* 16(1). 64–99. DOI: 10.1349/PS1.1537-0852.A.486.

Treis, Yvonne. forthcoming. *Ideophones in Kambaata (Cushitic): Grammar, Meaning and Use.*

Treis, Yvonne. in press. Reflexive constructions in Kambaata. In Katarzyna Janic, Nicoletta Puddu & Martin Haspelmath (eds.), *Reflexive Constructions in the World's Languages.* Berlin: Language Science Press.

Treis, Yvonne & Deginet Wotango Doyiso. 2019. "Issues and maize bread taste good when they're cool": Temperature terms and their metaphorical extensions in Kambaata (Cushitic). *Studies in African Linguistics* 48 (2). 225–266. https://doi.org/10.32473/sal.v48i2.118041.
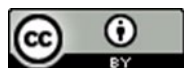
Yvonne Treis
LLACAN, UMR 8135 du CNRS
7, rue Guy Môquet - BP 8
94801 Villejuif, France
yvonne.treis@cnrs.fr

Hélène Vassiliadou, Francine Gerhard-Krait, Georgia Fotiadou
& Marie Lammert

Pseudo(-) in French and Greek: Categorization and approximation

Abstract: *Pseudo(-)* is generally considered to be a prefix or a prefixoid unit with an evaluative value of pragmatic approximation and negative polarity: *pseudo(-)*X presents a distortion with respect to a standard X. It then enters the paradigm of expressions responsible for infusing subjectivity into the discourse. The interpretative mechanisms in play exploit two dimensions, semantic and pragmatic, according to principles that will be explained in this paper. Furthermore, our contribution aims to highlight the semantic dimension in the interpretation of *pseudo(-)* both in Greek and French and to show how it exploits the semantic and categorial properties of the nouns that are postposed to it, leading to three main types of categorization (*pseudo(-)*X belongs to the category X but it is not a good exemplar, *pseudo(-)*X is clearly not an X, the categorization remains undecidable) and to several readings.

Keywords: categorization, approximation, disproximation, vagueness, evaluative morphology

1. Introduction

In this study, we examine *pseudo(-)* in French and in Greek, considered to be part of the paradigm of typical evaluative prefixes indicating "deviation from a standard value" as described, among others, in Štekauer, Salvador & Körtvélyessy (2012), Körtvélyessy (2015), Amiot & Stosic (2015, 2023) and Melissaropoulou (2015). Noticeably, there are no studies dedicated to *pseudo(-)* in French, except for the recent work of Van Goethem, Norde & Masini (2021/under review), while Greek scholars focused on the investigation of its morphological status (Giannoulopoulou 1997) and discussed primarily some of its aspects within the paradigm of diminutive verbs (Efthymiou 2017a, b, 2019).[1]

---

[1] See also Thillmann's study (2007) for German, Van Goethem & Norde (2020) for Dutch, Van Goethem, Norde & Masini (2021/under review) for eight European languages (Danish, Dutch, English, German, Swedish, French, Italian, Spanish) and Cappelle, Daugs & Hartmann (2023) for English.

In our turn, we aim to tease apart the various interpretative effects of *pseudo(-)* when used with nouns and to show that it always triggers categorization by signaling a gap between an individual or a class of *pseudo(-)*X and the category X itself. Three types of categorial relationships result: for example, a *pseudo(-)wife* out of context a) belongs to/is included in the category 'wife' but it is not a good exemplar (she does not correspond to social stereo-types), b) is clearly not a wife (officially speaking), c) the categorization remains undecid-able (the context does not offer enough information to decide on its status). Both the type of categorization triggered and the readings associated to it must be carefully established, as they do not always depend on a free choice of the speaker. As suggested in several recent studies (see i.a. Vassiliadou & Lammert (eds.) 2022; Mihatsch et al. (eds.) 2023), the dis-tinction between types of categorization as well as what falls under approximation is, in some cases, difficult to grasp, and objective criteria which, ideally, leave no room for doubt, are sometimes missing. For instance, glosses by *fake*, *similar to*, *alleged*, *so-called*, *quasi* and other devices that could help us decide which interpretation to choose, are not finally able to disambiguate the interpretations, because they convey the same semantic effects as *pseudo(-)* itself.

In order to untangle the skein of meaning effects of *pseudo(-)*X expressions and to grasp the semantic and pragmatic dimension in which meaning is built, we proceed as follows: first, we describe the relationship between categorization and approximation, mechanisms which we consider two sides of the same coin. We also plead for a semantic account of vagueness, and we discuss some morphological issues which also justify our methodologi-cal protocol (Section 2). In Section 3, we present the corpora we used for our investigation and our distributional data. Based on *pseudo(-)*'s general semantic information, shared by French and Greek, we then evaluate how it combines with the element X (=Noun) that follows it (Section 4). Within this perspective, three types of categorization are proposed. We specifically focus on the type of nouns that occupy the X-position (Section 5), since their semantic properties offer some objective criteria that may lead to an understanding of how the meaning of *pseudo(-)*X is specified. The main findings of this work as well as some open questions are discussed in the final part of the paper (Section 6).

2. Preliminary remarks on approximation and morphology

For the sake of functional comparison (which is often done in contrastive studies), we intentionally quasi-overlook the specific formal properties of the two formatives in French and Greek. Moreover, we consider that comparing a formative in its original language (even if it has travelled through history and languages; see Van Goethem, Norde & Masini 2021/under review) and in a target language is relevant for contrastive linguistics.

Despite the fact that French and Greek *pseudo(-)* differ in some respects as the presence of two forms and the freer distribution of the Greek marker (see Section 2.2), they both share the same semantic core meaning. They enter the domain of evaluative morphology by providing appreciation of the existing conceptual categories, not only by focusing on diminution and depreciation, but also by considering the intersection between categorization/approximation and evaluative morphology (see Masini, Norde & Van Goethem 2023, Introduction of this issue, for references). Thus, we put forward that, in all its uses, both in Greek and French, *pseudo(-)* is clearly a hedge[2] in Lakoff's original sense (1973), that is an expression whose role it is to make a categorial belonging either clearer (taxonomic reading) or less clear (evaluation of the degree of exemplarity within a category, i.e. intra-category approximation, and fuzzy reading). Nevertheless, that does not imply that *pseudo(-)*X is a vague expression, as we will show below.

2.1 Clear and approximate categorization: Two sides of the same coin

We will not investigate the terminological imbroglio that closely affects *approximation*, *imprecision*, *vagueness* and *categorization*: approximation is often correlated with vagueness, vagueness or fuzziness are correlated with pragmatics, the boundaries of vagueness and indeterminacy are fuzzy in turn (Kleiber 1987). Moreover, it seems that, in recent years, scholars tend to overlook the fact that vagueness is also an affair of semantics and prefer to associate it quasi-exclusively with pragmatics and context-dependencies. Devos (2003: 122–123) argues that vagueness is "primarily a semantic phenomenon, and not a pragmatic one, as vagueness cannot always be imputed to language users". In fact, we deal, above all, with a language phenomenon and not an extra-linguistic one; semantic

---

[2] The concept of *hedge* is used in this paper as a "comparative concept" (Haspelmath 2010) applicable to Greek and French (and probably to other languages which use different sorts of hedging devices).

vagueness is, thus, conceived as "an intrinsic uncertainty with regard to the application of a word to a denotatum" (Devos 2003: 123). Following Devos, we will show, for instance, how objectively vague predicates (like nouns designating affects) play a role in the interpretation of *pseudo(-)*X.

According to psycho-/neurolinguistics, the principle of categorization itself functions by means of relating elements (Kahlaoui et al. 2010). This is also the case for approximation and comparison. In other words, both taxonomic and approximative categorization relate some element to a category. Assigning a category to a referent and judging the adequacy of its categorial membership can in some cases be difficult. The categorization at stake for a given statement, in terms of clear, approximate or vague membership (see Section 4), can sometimes vary considerably depending on the interpreter or the descriptor (see, among others, Gerhard-Krait & Vassiliadou 2014, 2017; Gerhard-Krait & Zerva 2023; Vassiliadou et al. 2023). The reasons that may explain this difficulty in apprehending the boundaries of certain categories are of various kinds: inherent semantic indeterminacy, vague predicates, speaker's intention and multiplicity of pragmatic values, to name a few.

Having established that our starting point is to avoid opposing categorization and approximation, we next show that inconsistency runs through morphological marking as well.

2.2 Morphological and syntactic status of *pseudo(-)*

Pseudo(-)'s morphological status as a pejorative, depreciative prefix or prefixoid (see among others Quirk et al. 1985; Van Goethem & Norde 2020) or as a bound morpheme entering neo-classical compositions (Baeskow 2004) or even as a "semi-word" ("semiparola" in Scalise 1990) can be discussed in length. Most cases, where *pseudo(-)* (and other similar items) appear, are considered in the literature as cases of occasional composition (syntactic constructions),[3] i.e. words constructed in discourse (Corbin 1987; Lieber 2004).

Pseudo(-), as far as French is concerned, can indeed be seen as an initial combining form which appears in nominal and adjectival units recorded in dictionaries, notably in specialized lexicons and terminologies, as in *pseudonyme* 'pseudonym', *pseudo-science* 'pseudo-science', *pseudo-kyste* 'pseudo-cyst', *pseudo-tourelle* 'pseudo-turret', *pseudo-bulbaire*

---

[3] Our aim is not to study when a word stops being an "occasional composition" and becomes a "real word". Literature on neology tries to answer at this very controversial question by using many different criteria like frequency, autonomy, semantic and formal stability, fixation, etc.

'pseudobulbar', *pseudo-byzantin* 'pseudo-byzantine', and so on. Given the number of oc-currences of *pseudo(-)*X which are not intended to form lexicalized units (which is one of the properties of evaluative morphemes), it could also be seen as a simple modifier of a noun or an adjective in French assuming the status of a "debonding unit" (Norde 2009: 186). Indeed, there are many occurrences in French where users proceed to agreement between *pseudo(-)* and X (like adjectives do): *pseudos qualités* ('false qualities'), *pseudos artistes* ('false artists') among others[4].

Greek data[5] can be tricky because *pseudo(-)* may surface with two forms: *ψευδο(-)* psevðo(-) and *ψευτο(-)* psefto(-) for which the literature has not reached definite conclu-sions concerning their categorization and does not uniformly agree on whether *psevðo(-)* and *psefto(-)* should be considered as allomorphs or not. This is the case, for instance, for Giannoulopoulou (1997), who considers *psevðo-* as a confix (i.e., neither stem nor affix) and *psefto-* as the first element of compound words. Her main argument is that *psevðo-* (from *ψευδ(ής)* 'lying, false') conserves in all cases the meaning of 'artificial, fake' and re-jects the evaluative one 'vile, junk' (Giannoulopoulou 1997: 120), leaving *psefto-* to endorse the evaluative meaning of depreciation. She also claims that *psevðo-* can be combined ex-clusively with nouns (which is not the case, even if nouns are indeed more frequent; see Section 3). If this statement was true, how could we explain their massive coexistence with the same 'base' and the same meaning (*ψευτο/ψευδο-επιστήμη* psefto/psevðo-epistími 'pseudo-science', *ψευτο/ψευδοδίλημμα* psefto/psevðoðílima 'pseudodilemma', *ψευτο/ψευδοανάγκη* psefto/psevðoanàgi 'pseudoneed', etc.) rather than free variation in speakers' usage or the well-established differentiation in Greek between ð used in more literary reg-isters and τ in more familiar ones? One can also argue that as Greek dictionaries register *psevð-* and *pseft-* under two different entries, they are not allomorphs. It is well known that this is not a convincing argument as lexicographers do not distinguish between polysemy and monosemy based on strict morphological criteria. Finally, a recent study (Anastassiadi-Symeonidi & Fliatouras 2019: 26) explains why Modern Greek allows the presence of both

---

[4] See also some cases where *pseudo(-)*'s scope is over a noun phrase (NP) (Van Goethem, Norde & Masini (2021/under review).

[5] We wish to express our gratitude towards our anonymous reviewers who encouraged us to further develop some aspects neglected in the first version of this paper. Many thanks also to Anna Anastassiadi-Symeonidi with whom we had extended discussions that helped us better understand many issues raised for Greek.

psefto(-) and *psevðo(-)* with the same 'base' by showing that there is a clear continuum from [+learned] to [−learned], the norm (everyday vocabulary) being in the center. The norm corresponds to a two-zone overlap (formal and familiar registers) including thus all permitted combinations within the Greek lexicon. Moreover, Ralli (2005: 169; see also Ralli 2004) considers this kind of formations as "pseudo-compounds": the presence of the vowel *o* in *psevðo(-)*/*psefto(-)* is not the typical compounding marker in Greek, but the last segment of the "affixoid". We can also find the characterization of "fossilized diminutive" referring to a very productive paradigm (*micro(-)* 'small', *psilo(-)* 'thin, a bit', etc.) in which the items are not "proper diminutives" (Xydopoulos 2009: 404).[6]

The aim of this paper is not to conduct a diachronic study (see Babiniotis 1969), but we can outline here some observations retrieved from Fotiadou et al. (forthcoming). *Psevð-* and *pseft-* are to be examined probably in the times of Puristic Greek (*Katharevousa*), but a quick glance at TLG corpus data (10th AD onwards; theological/historical texts) shows the possible existence of two derivational paths for *pseudo(-)*:

1. *pseud(o)-* complement/argument of noun as in *ψευδάγγελος* psevðàgelos, lit. pseudo-angel 'who announces lies', or in *ψευδολόγος* psevðológos 'who tells lies, liar'. These formations can be glossed by a corresponding phrase.

2. *pseudo(-)* specifier/adjunct in cases which we cannot easily paraphrase by using corresponding phrases: *ψευτοδουλεύω* pseftoðoulévo, lit. pseudowork, does not mean 'work in a false/artificial way' but 'work a little' or 'pretend to work' (see Efthymiou 2017b: 5–11).

The first path is still at work in Modern Greek and even if rare, it is not negligible: we encounter it with lexicalized units as *ψευδομάρτυρας* psevðomártiras 'false witness, perjurer' and within the scientific vocabulary as *ψευδοκύστη* psevðokísti 'pseudocyst'. The second path, the dominant one in our corpus data and in general, is very prolific/productive, in line with the affixoids' Modern Greek paradigm. Even if the meaning of 'fake' seems transparent, we do not claim that *pseudo(-)* in *pseudo(-)*X has fully retained its original meaning; it is on the contrary an underspecified item that, as *psilo(-)* 'thin, a bit', "has undergone bleaching […], lost some of its semantic features and become more and more

---

[6] For diminutive formation in Greek, see Melissaropoulou & Ralli (2020).

abstract" (Xydopoulos 2009: 400). The transition from 'fake, artificial' towards 'of a lower quality' is easy to grasp. It is interesting to note here that in some cases, when the process of grammaticalization is not fully achieved, ambiguities remain. Evidence from Sarantakos blog show the disagreement between the lexicographer's position and the comments uploaded by many users of the internet on the interpretation of *ψευτοπονηράκιας*[7] pseftoponirákias ('pseudo-cunning'):

- Not a pseudo-cunning but a liar and a cunning (for the lexicographer).
- Pseudo-cunning is a person who pretends or is trying to persuade that he is clever though he is not (for internauts). (sarantakos.wordpress.com/2009/04/02/nopseudo-ponirakias)

For both French and Greek, we consider all the possible orthographic marking variations of *pseudo(-)* (welding, hyphen, absence of hyphen, in brackets) insofar, as they are not reliable criteria to differentiate what would be lexicalized units. This is supported by the fact that a lexicalized unit can appear with or without a hyphen (*pseudo-rhumatisme/pseudorhumatisme* 'pseudo(-)rheumatism', *ψευδο-ιστός/ψευδοιστός* psevðo(-)istós 'pseudo(-)tissue', for example), just as free uses of *pseudo(-)* can appear with or without a hyphen (*pseudo-mari/pseudo mari* 'pseudo(-)husband', *ψευτοόραμα/ψευτο-όραμα* psefto(-)órama 'pseudo(-)vision').[8] Moreover, dictionaries register both hapaxes and occasionalisms, thus they do not guarantee the denomination status of the unit. Queries in corpora also display discrepancy regarding the presence or absence of the hyphen.[9]

While we do not assimilate prefixal and syntactic mechanisms,[10] we rely on the fact that for the constructed lexicon, there is a prior fixation of meaning, whereas for free sequences,

---

[7] Note that the ambiguity concerns even *psefto(-)*, which is supposed to specialize in depreciation.

[8] Nevertheless, welding is a sign of lexicalization in French only as far as lexicalized units are concerned.

[9] In the literary French corpus Frantext, the tendency is to hyphenate almost all *pseudo(-)*X. Statistics on Wortschatz and frTenTen17 are difficult to present here due to the high frequency of *pseudo(-)* as the truncation of *pseudonym*. Van Goethem, Norde & Masini (2021/under review) present in their paper token frequencies of French *pseudo*'s construction types as per 1000 tokens extracted from TenTen web corpus. They found 393 neo-classical compounds, 192 clippings (i.e., *pseudo* meaning 'nickname') and 415 native formations. In Greek, things are a little bit different as there are almost the same number of tokens with and without hyphens, but also a great number of *pseudo(-)* in brackets: *(ψευδο)ιστορία* (psevðo)istoría '(pseudo)history', *(ψευδο)επανάσταση* (psevðo)epanástasi '(pseudo)revolution' (see Fotiadou et al. forthcoming).

[10] As they do not belong to the same functional domains, they do not follow the same rules of combination and they do not produce the same semantic and pragmatic effects.

the semantic value primarily obeys the semantic potential of the nouns associated to the *pseudo(-)*X sequence. Thus, there is always a computation of meaning to be made in context. For example, *pseudo-tissu* 'pseudo-tissue' as a lexical unit designates something that is not a tissue. We could imagine that a free use of *pseudo(-)* combined with *tissu* means a tissue of bad quality.

Hence, *pseudo(-)* can be seen as an element at the crossroads of word construction and syntactic construction. Whether it is free or bound, it produces the same interpretative effects and, for this reason, we do not take into account considerations related to its status as a prefix or an autonomous element, as this is not relevant to our semantic approach. In all its uses, both in Greek and French, *pseudo(-)* is clearly a hedge in Lakoff's original sense (1973), as stated from the outset.

In the following sections, we seek to understand the role that the element X plays in the different interpretative schemes when X is a noun.

3. *Pseudo(-)* in French and in Greek: Corpus and distribution

Taking under consideration the above, we focus on the semantics of *pseudo(-)* in all its formal configurations with respect to the type of nouns appearing in the position X, but we also examine to what extent a pragmatic dimension is called upon. For the purposes of this paper, we opted for a qualitative analysis of the examples given, even if some quantitative data are also provided to justify our choices regarding the phenomena affecting *pseudo(-)*'s uses.

3.1 Corpus

The data extraction both in Greek and French was motivated by the need to cover as many registers as possible. For French, we consulted the following three corpora: first, the literature database *Frantext* gave back 1,441 tokens of *pseudo* from 1610 to 2019. Secondly, the *fra_mixed_2012* corpus, available on the *Leipzig Corpora Collection – Wortschatz,* composed of various texts (from newspapers, webcrawls, etc.), returned 9,582 tokens of *pseudo* and *pseudos,* from which we retrieved the first 200 tokens of each form. This corpus also

allowed us to extract a list of 1,370 words composed on the scheme pseudo* or pseudo-*[11]. Thirdly, we found 121,673 tokens of *pseudo* (excluding *pseudonyme*) in the French Web 2017 Corpus (*frTenTen17*), from which we retrieved a random sample of 400 tokens and a list of 997 words composed on the scheme pseudo* or pseudo-*. As expected, the results in these corpora and databases contained some noise, since *pseudo(-)* can be used to designate, by truncation, a *pseudonym*, which was manually eliminated and relevant tokens were not considered in the analysis.

For Greek, we also consulted three corpora: in the first place, we used the *Thesaurus Linguae Graecae* (TLG) to register the initial uses of *pseudo(-)*, diachronically speaking, as already mentioned in 2.2 above, but we do not further exploit in this paper the tokens found in this base (see Fotiadou et al. forthcoming). For Modern Greek, we used the Greek Web 2014 (elTenTen14) corpus, where our searches were limited to creations with *ψευδ-* psevð- (n=556) or *ψευτ-* pseft- (n=665). We finally exploited some examples from a broadly used corpus of written speech (various genres), the Hellenic National Corpus of Institute for Languages and Speech Processing, where searches for *ψευδ-* psevð- (n=64) or *ψευτ-* pseft- (n=40) returned very few example formations.

3.2 Distribution

As far as the distribution of X (in *pseudo(-)*X) is concerned, snippets extracted from the corpora show a clear-cut tendency for *pseudo(-)* to be attached to nouns in both languages. Nevertheless, we cannot provide reliable statistics for French (especially for frTenTen17 and Wortschatz) due to the many incorrect annotations. Adjectives related to specialized domains (medicine, architecture, art, terms that designate an era, a period) are also quite frequent[12] (*pseudo-bulbaire* 'pseudo-bulbar', *pseudo historique* 'pseudo-historical', *pseudo-gothique* 'pseudo-gothic', *pseudo-classique* 'pseudo-classical'). In the literature corpus Frantext, *pseudo(-)* is used with nouns almost twice as much as with adjectives (Fig. 1). In French, combinations with adverbs could be conceivable but are not attested in our

---

[11] Strings used to extract the hits.

[12] For TenTen web corpora, see the statistics in Van Goethem, Norde & Masini (2021/under review): high proportion of clippings (19,2%), majority of nouns (32%) followed by adjectives (ca. 12%).

corpora (nor accepted by native speakers of French[13]), except for 12 tokens of *pseudo-aléatoirement* derived from the lexicalized adjective *pseudo-aléatoire* 'pseudo-randomized'.
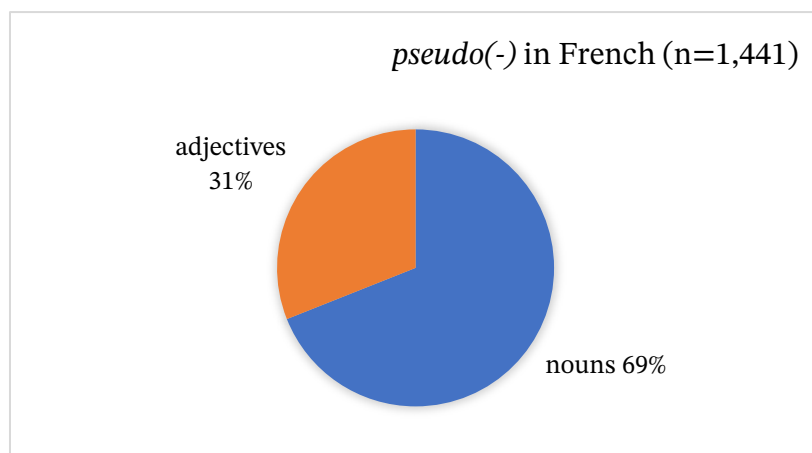
Fig. 1: Distribution of *pseudo(-)* in French in Frantext

On the other hand, the distribution of *pseudo(-)* (noticeably *pseft-*) in Greek is very rich (see Fig. 2). It can combine with nouns (*ψευτο επιχειρήσεις* psefto epixirísis 'pseudo enterprises', *ψευδο αδιακρισία* psevðo aðiakrisía 'pseudo indiscretion/intrusiveness'), adjectives (as in *ψευδοδημοκρατικός* psevðoðimokratikós 'pseudodemocratic', *ψευδογοτθικός* psevðo-γotθikós 'pseudogothic', *ψευδορομαντικός* psevðoromantikós 'pseudoromantic', *ψευδο-βυζαντινός* psevðo-vizantinós 'pseudo-byzantine'), verbs[14] (*ψευτοπαραπονιέμαι* pseftoparaponiéme 'to pretend to complain/to complain a bit', *ψευτογκρινιάζω* pseftogriniázo 'to grumble a bit', *ψευδο αρνούμαι* psevðo arnúme 'pseudo deny'), adverbs (*γράφω ψευτορομαντικά* gráfo pseftoromandiká 'to write in a pseudo romantic manner', *ψευτο ήρεμα* psefto írema 'in a pseudo cool way'), interjections (*Άσε τα ψευτοάχ σου τώρα* ase ta pseftoáx su tora 'leave now your false suffering') and pronouns (*ψευτοεγώ*, pseftoegó 'pseudo-I').

---

[13] We used a small sample of informants (n=10), who provided us grammaticality judgment regarding the well-formedness of a set of examples with *pseudo(-)* which we no further discuss.

[14] All data were cleared and annotated. *Psevðo(-)* is by far less compatible with verbs than *psefto(-)*. We found few tokens of verbs such as *ψευδοσυντηρώ* psevðosindiró 'pseudo conserve' or *ψευδογεμίζω* psevðojemízo 'pseudo fill' where *psevðo(-)* has an evaluative value.
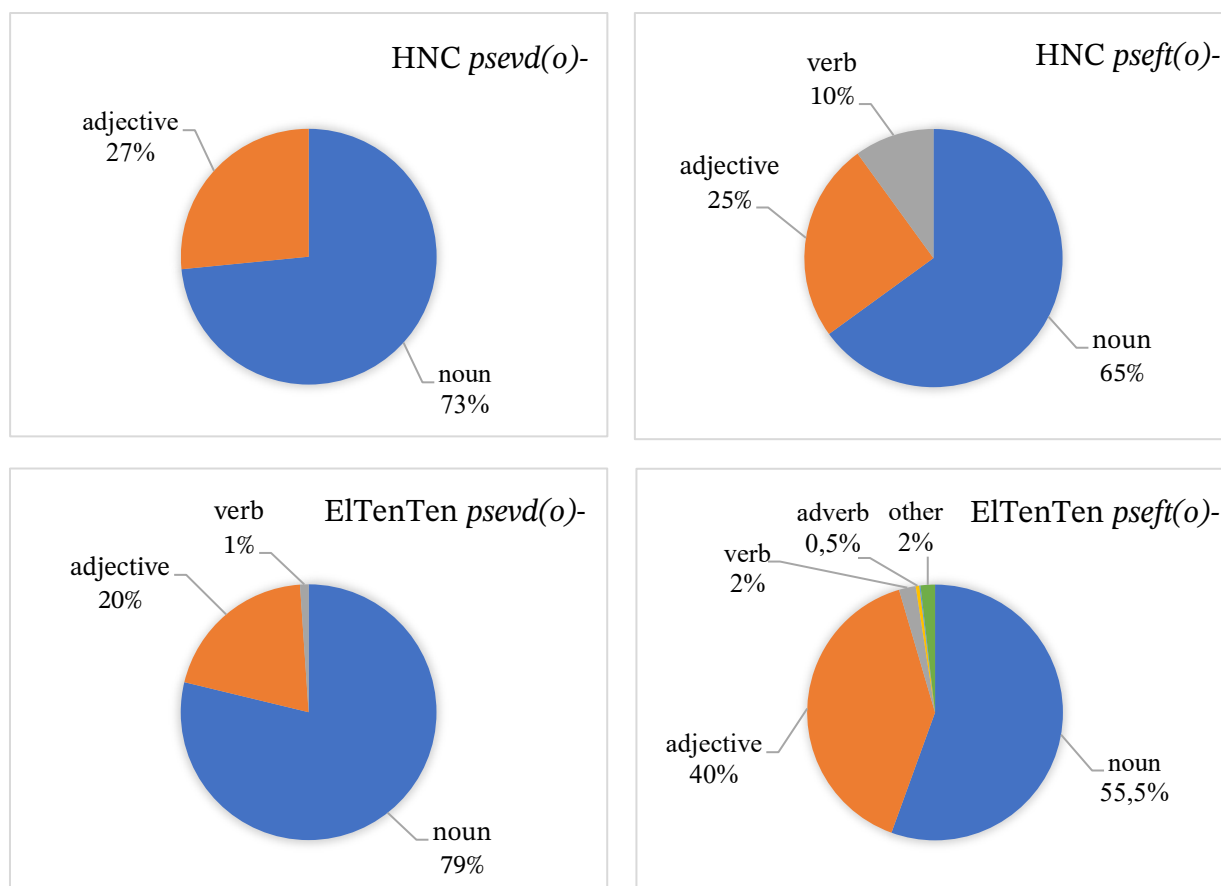
Fig. 2: Distribution of *pseudo(-)* in corpora of Modern Greek

Combinations with verbs are a very interesting phenomenon that needs to be thoroughly described. This kind of combination, even if conceivable in French, does not seem to occur. For instance, if one says in French something like *Je pseudo-dors* 'I pseudo-sleep', we may easily understand the implied meaning, even if it is not attested in naturally occurring data. On the other hand, when one says in Greek *ψευτοκοιμάμαι* pseftokimáme 'I pseudo-sleep', we understand either something like 'I am sleeping but in a non-qualitative way' (because my sleep is disturbed or because I know that I have to wake up soon, etc.) or 'I am dozing' ('I do not really sleep', the speaker is at the beginning of the activity of sleeping).[15] In (1), *(ψευτο)κολυμπήσαμε* (psefto)kolimbísame 'we pseudo-swam' means either that the speaker and his friends did swim, but not enough, not for a long time or that they just floated (what they did cannot be called swimming strictly speaking). Accordingly, *(ψευτο)ψαρέψει* (psefto)psarépsi '(had) pseudo-fished' signifies either that they went fishing but they caught no fish (or just a few) or that the activity cannot be called *fishing* because the

---

[15] Verbal aspect is also to be considered.

appropriate equipment was missing. In any case, the *pseudo-activity* took place, but its quality is depreciated:

(1)   *ξέρω, είναι όμως εκπληκτική κι σ'ένα παρθένο μέρος! Εμένα μου θύμισε τα Φαλάσαρνα της Κρήτης! Πέρσι το καλοκαίρι **(ψευτο)κολυμπήσαμε** εκεί [οικογενειακώς] σε επίσκεψη μας στην περιοχή όπου η παρέα είχε **(ψευτο)ψαρέψει** κιόλας.* (epanenosi.com)

kséro, íne ómos ekpliktikí ki s'éna parθéno méros! Eména mu θímise ta Falásarna tis Krítis! Pérsi to kalokéri **(psefto)kolimbísame** eki (ikoγeniakós) se epískepsi mas stin perioxí ópu i paréa íxe **(psefto)psarépsi** kiólas.

'I know, but it's amazing even in a virgin place! It reminded me of Falassarna in Crete! Last summer we **(pseudo)swam** there [as a family] on a visit to the area where the group had even **(pseudo)fished**.'

As Efthymiou (2017b: 5) pointed out, in the Greek verbal domain,

psefto- appears frequently with verbs denoting activity [...] and indicates that the process in question is performed with less effort than expected. In addition to expressing the speaker's attitude towards the event in question, verb formations with *psefto-* can also express event internal pluractionality [...].

4. What kind of categorization for *pseudo(-)*?

*Pseudo(-)*X, as already suggested (Section 2), always implies categorization; it presupposes a comparison between what it designates and an expression X, and it consists of an evaluation of the categorial adequacy between what *pseudo(-)*X designates and the category X itself. This evaluation is negatively polarized, i.e., it focuses on features of difference and deviation from the category (see the term of *disproximation*; Cappelle, Daugs & Hartmann 2023). Thus, the referent of *pseudo(-)*X (where *X* is a noun) deviates in one way or another from X.

In what follows, we first present the three general types of categorization triggered by *pseudo(-)* (4.1–4.3)[16] and we next associate them to the types of readings they receive along with the X's type noun which participates in the expression (Section 5).

---

[16] In this section, we present one of the possible interpretations of *pseudo(-)*X out of context: the same examples are also examined in context (Section 5) in order to highlight the different readings.

4.1 *Pseudo(-)*X looks like an X but is not an X

*Pseudo(-)*X denotes another category, notably in many denominations (see *pseudo-kyste* ψευδοκῦστη psevðokísti 'pseudocyst' and commented examples (2), (5), (6)) or in occasional designations[17] (as ψευδομετάξι psevðometáksi 'pseudo silk', *pseudo-havane* 'pseudo-Havana cigar', *pseudo-banquise* 'pseudo-ice pack', *pseudo-Mozart* 'pseudo-Mozart'). Artefacts are the best candidates as *pseudo(-)* denotes forgery or counterfeiting (*pseudo-iPad*). There is always at least one objective (i.e., semantic) feature which allows us to oppose X to *pseudo(-)*X: the deviations concern the semantic properties of X in such a way that *pseudo(-)*X does not satisfy the conditions of applicability of the denomination X as in (2)–(6):

(2)   A *pseudo-membrane* 'pseudo-membrane' looks like a membrane (is a membrane-like secretion) but is not a membrane. These are two different categories.

(3)   *Pseudo-Mozart* is not Mozart, but another composer who plays/composes like/in the style of Mozart.

(4)   A *pseudo-gestation* (lit. pseudo-pregnancy) 'phantom pregnancy' shows symptoms of gestation, but there is no baby in the end.

(5)   Ψευδο-ιστός psevðoistós 'pseudo-tissue' is an artificial tissue made in a lab.

(6)   A *pseudo-mot*/ψευδολέξη psevðoléksi 'pseudo-word' has the form of a word but it lacks semantic content (it is not a linguistic sign; see speech therapist experiments: it follows the phonotactic rules of a language and is different from *non-words*, where specific violations are applied).[18]

In these commented examples, *pseudo(-)*X is objectively a 'no X', it is not part of the semantic extension of X. There is no subjective devaluation, but there are clear cases of non-inclusion in the category, although some similarities are shared (e.g., shape). The possible following glosses for these cases of non-inclusion leading to a clear interpretation of the examples can be applied:

---

[17] The same referent can receive two linguistic realizations: 'church' is a denomination vs. 'the place of worship for Catholics' is a designation (Kleiber 2001).

[18] See the term of "kin-categorization" (i.e., the creation of a new closely related category) used by Masini & Micheli (2020).

- *Pseudo(-)*X looks like an X but is not an X.
- It is a fake X (as in real fake, strictly speaking).
- It is an imitation of X, an artificial X, a forgery.
- It pretends to be X, but is not, it passes itself off as X.

When denominations are formed, *pseudo(-)*X can create in turn a category, X and *pseudo(-)*X can then be co-hyponyms. This is the case of 'pseudo cyst', which is an approved term to denote a type of pathology.

4.2 *Pseudo(-)*X is included in the category X but is qualitatively depreciated

We now proceed with the discussion of a second type of categorization with *pseudo(-)*, which concerns what is known in the literature as *approximation* (see among others Mihatsch 2009). The approximation in question is mainly a subjective depreciative evaluation and, as such, relevant to pragmatic vagueness. *Pseudo(-)*X is objectively X but the speaker adds subjective judgement, and thus modality: if a real, actual *team supporter* is qualified as *a pseudo-supporter*, this means that the speaker judges that he is not a good exemplar of the category, he does not deserve to be called a *supporter*, as in (7); similar cases are included in (8)–(11):

(7)   A *pseudo-supporteur* 'pseudo-supporter' denotes a team supporter who is negatively evaluated by the speaker.

(8)   A *pseudo-mari* 'pseudo-husband' can be officially a husband but a failing one.

(9)   A *pseudo-enquête* 'pseudo-investigation' is an actual investigation (conducted by police officers or journalists, for instance), but the speaker considers it to have flaws, because it is based on approximations, innuendos, false allegations, etc.

(10)   *Ψευδο-δημοσιογράφος* psevðo-ðimosiográfos 'pseudo-journalist' can be an official member of the order of journalists (with accreditations, etc.) but a bad one, one who is doing bad journalism.

(11)   *Ψευτο συγγραφέας* psefto sigraféas 'pseudo writer/author' can be an acknowledged author, but depreciated by the speaker.

Possible glosses for the commented examples (7)–(11) are provided:

- *Pseudo(-)*X is a non-exemplary X, it is a member considered as (negatively) deviant.
- *Pseudo(-)*X is a bad X, an X of poor quality, an X that does not fulfil its role, etc.

All these cases belong to the negative evaluation of a member of a category. This is by far what we would call *subjective (modal) categorization. Pseudo(-)*X is a non-prototypical or non-stereotypical X. It draws its value essentially from connotations, from additional features of X, or of a speaker's judgment. It behaves as a "derogatory marker" (Xydopoulos 2009) similarly to *psilo(-). Pseudo(-)* exploits the vagueness not from a semantic point of view, but from the evaluative judgment on the connotations attached to X or compared to a prototype.

4.3 Vagueness: Undecidable Categorization

In this third case, we are dealing with instances of undecidable categorization, i.e., occurrences where there is no more reason to think that *pseudo(-)*X is included in the category X than the opposite. Semantic vagueness meets here subjective vagueness: as *pseudo(-)* exploits the existence of borderline cases and underlines the negative side of X, it questions the very categorization of X. Contrary to what we exposed above, in undecidable categorization we lack objective features to oppose X to *pseudo(-)*X. It is then necessarily the speaker who decides whether *pseudo(-)*X belongs to the category of X or not.

This type of *pseudo(-)*X typically concerns controversial predicates about which we can argue at length: for instance, "what is love?", "what is poetry?" "who can be called an artist or an intelligent person, an expert, a seducer, etc.?" vs. "what is a tissue, a membrane, a Havana cigar, an iPad, etc.?" for which there are no objective borderline cases, as in (12) and (13):

(12)   *Pseudo-liberté* 'pseudo-freedom', *pseudo-amour* 'pseudo-love'

(13)   *Ψευδοποίηση* psevðopíisi 'pseudopoetry', *ψευτοδίλημμα* pseftoðílima 'pseudodilemma'

Possible gloss for these cases:

  – *Pseudo(-)*X claims/is claimed to be X but could just as well be non-X.

We next present the distribution of noun types that enhance specific types of the three categorizations of *pseudo(-)X* and the readings they favor.

5. Types of nouns, types of categorization and types of readings

In this section, based on well-established nominal typologies (see Flaux & Van de Velde 2000; Huyghe 2015), we propose an interpretative prediction model for *pseudo(-)*X

depending both on the semantic and categorizing potentials of the X-noun and the semantic value associated to *pseudo(-)*.

5.1 One type of categorization, one type of reading

5.1.1 Categorial exclusion reading

There are only a few nouns that imply a single reading of *pseudo(-)X*, namely that *pseudo(-)X* does not belong to the X category.[19] *Pseudo(-)X* is a copy, an imitation (even if there may be also a depreciative value because of the counterfeit character). The category of X is clearly established. The nouns belonging to this class fall under the pattern seen in Section 4.1: *pseudo(-)X* looks like X but is not an X.[20]

Artefacts (especially with a brand name like *iPad*, *Havana*, *Panama*, *Nike*, etc.) tend to be the best candidates for categorial exclusion readings, as illustrated in (14)–(18):

(14) *Le bout du havane, ou **pseudo-havane**, s'illumina et resplendit.* (Arnoux, *Double chance*, 1958)

'The end of the Havana cigar, or **pseudo-Havana**, lit up and glowed.'

(15) *Roubaud (…) arbore un **pseudo-panama** sous lequel il se croit très chic.* (Colette, *École*, 1900)

'Roubaud (…) wears a **pseudo-panama** under which he thinks he is very chic.'

(16) *οι μόνοι δυτικοί ανάμεσα σε καμιά τετρακοσαριά έγχρωμους συνταξιδιώτες με **ψευτο-nike**, φθαρμένα παπούτσια και κουρασμένα χαμόγελα.* (koel.gr)

i móni ðitikí anámesa se kamiá tetrakosariá éxromus sintaksiðiótes me **psefto-nike**, fθarména papútsia ke kurazména xamójela.

'the only westerners among about four hundred colored travelers with **pseudo-Nike**, worn out shoes and tired smiles.'

---

[19] Some authors use the term *privative*. If by *privative* we mean that *pseudo(-)* deprives X from the category, this terminology can be also applied to our examples. But semantically speaking, *pseudo(-)* is also an evaluative marker. In this paper, we consider it as a hedge as stated in Section 2.

[20] The combination of *pseudo(-)* + adjective indicating historical periods or architectural styles is used to always signal categorial inadequacy. For instance, in the following example, the style of Saint Augustine's church in Paris has a byzantine allure, it resembles byzantine art without being of it: *(L'église saint Augustin) de style **pseudo-byzantin** s'inscrit dans les travaux d'Haussmann sous Napoléon III (…)* '(The church of Saint Augustine) of **pseudo-Byzantine** style is part of the work of Haussmann under Napoleon III' (https://www.europexplo.com/visite-de-leglise-saint-augustin).

(17)   *συνεχίζεις, πωλώντας τα (ψευδο-)μεταξωτά σου (και) σε πάμφθηνη, μάλιστα, τιμή.* (blockspot.cz)

sinexízis polóndas ta **(psevðo-)metaksotá** su (ke) se pámfθini, málista, timí.

'you continue, selling your **(pseudo-)silks** at a bargain price.'

(18)   *Στον όροφο υπάρχουν και ψευτοπαράθυρα ιδίως στα αρχοντικά, καθαρά για λόγους διακόσμησης.* (pblogs.gr)

Ston órofo ipárxun ke **pseftoparáθira** iðíos sta arxondiká, kaθará ja lóγus ðiakózmisis.

'On the floor there are also **pseudo windows**, especially in the mansions, purely for decorative reasons.'

It is interesting to note that in (14) the speaker is shown unable to decide whether the referent *is* or *is not* a real Havana cigar, but (s)he knows that there is a semantic, a categorial difference between X and *pseudo(-)*X ("observational vagueness" in Raffman 2011).

Proper names, other than brand names, especially artists' or politicians' proper names are also good candidates for categorial exclusion reading, as shown in (19) and (20):

(19)   [...] *le **pseudo-Mozart** signé Strauss, a encore, pour beaucoup, l'air* [...] *pas sérieux et légèrement pervers* [...] (Le Nouvel Observateur, 2 February 1976).

'the **pseudo-Mozart** signed Strauss, still looks [...] not serious and slightly perverse [...]'

(20)   *ο Παπανδρέου βγαίνει σαν νέος (ψευτο-)Μεταξάς ενός νέου Γενναίου ΟΧΙ και λέει: Εγώ ήμουν αποφασισμένος να το πάω στα άκρα* [...] (mypharm.gr)

o Papanðréu vγéni san néos **(psefto-)Metaksás** enós néu jenéu OXI ke léi: Eγó ímun apofasizménos na to páo sta ákra [...]

'Papandreou comes out as the new **(pseudo-)Metaxas** of a new Brave NO and says: I was determined to take it to the extreme [...]'

Noticeably, the referential uniqueness of proper names in (19) and (20) leaves no doubt about the inadequacy of the application of the proper name to the referent.

5.1.2 Devaluative or subjective reading

Due to their very wide semantic extension and their very low intension,[21] general nouns (including some nouns called in the French literature *noms d'idéalités* 'ideality nouns';

---

[21] "Intension" indicates the internal content of a term or concept that constitutes its formal definition, while "extension" indicates its range of applicability by naming the particular objects that it denotes.

Flaux & Stosic 2015; Stosic & Flaux 2021) and some shell nouns (Schmid 2000), when associated with *pseudo(-)* drive the categorization towards category membership and the reading towards a negative qualitative evaluation (devalued belonging). In other words, general nouns are semantically underspecified but they are not necessarily vague predicates. Despite their restricted semantic intension, they draw the boundaries of the category in a clear way.

If X is considered an idea, a problem, an element or an argument, it necessarily belongs to the attributed category of ideas, problems, etc., and *pseudo(-)* can only make a negative axiological evaluation. To put it briefly, what else can an idea be, without being an idea? As shown in (21) below, a 'pseudo-idea' is still an idea, but maybe bad, confusing, or difficult to apply. Similarly, a 'pseudo-problem' (22)–(23) or a 'pseudo argument' (24) and even a 'pseudo-dialogue' (25) cannot be conceived otherwise:

(21)   *L'idée d'une abolition de tout est donc destructive d'elle-même, inconcevable ; c'est une* **pseudo-idée***, un mirage de représentation* (Bergson, *Deux sources*, 1932)

         'The idea of an abolition of everything is thus destructive of itself, inconceivable; it is a **pseudo-idea**, a mirage of representation'

(22)   *Il y a là un* **pseudo-problème** *; si la conversion est pensée, c'est parce que l'expérience nous en fournit des exemples ; il est impossible de se poser ici un problème à proprement parler critique.* (Marcel, *Journal*, 1914)

         'There is a **pseudo-problem** here; if conversion is thought of, it is because experience provides us with examples of it; it is impossible to pose here a problem properly speaking critical.'

(23)   *Ακόμα, η αξιολόγηση των δράσεων μιας επιχείρησης με ωφελιμιστικά κριτήρια οδηγεί σε αναπάντητα* **(ψευδο)προβλήματα***, του τύπου αν είναι ηθικά θεμιτό ή όχι μια επιχείρηση να απολύσει 1.000 εργάτες σε μια πόλη της χώρας της και να εγκατασταθεί (με σκοπό τη μείωση του κόστους) σε μια άλλη χώρα προσφέροντας εργασία σε 5.000 ή και περισσότερους εργάτες που ζουν* [...] (ethics.gr)

         Akóma, I aksiolójisi ton ðráseon mias epixírisis me ofelimistiká kritíria oðijí se anapándita **(psevðo)provlímata**, tu típu an íne iθiká θemitó i óxi mia epixírisi na apolísi 1.000 erɣátes se mia póli tis xóras tis ke na eɣatastaθí (me skopó ti míosi tu kóstus) se mia áli xóra prosférondas erɣasía se 5.000 i ke perisóterus erɣátes pu zun [...]

         'Still, the evaluation of the actions of an enterprise with utilitarian criteria leads to unanswered **(pseudo)problems**, such as whether or not it is morally legitimate for an enterprise to lay off 1,000 workers in one city of its country and move (in order to reduce costs) to another country offering work to 5,000 or more workers living in [...]'

(24)   *Comme ces **pseudo arguments** qui sciemment mettent en perspectives le nucléaire avec des champs d'éoliennes.* (rss.feedsportal.com)

'Like those **pseudo arguments** that knowingly put nuclear power in perspective with wind farms.'

(25)   *Με απάτη έγιναν **ψευτο-"διάλογοι"** με τις άμεσα πληττόμενες κοινωνικές ομάδες, όταν οι βασικές αποφάσεις ήταν ήδη ειλημμένες.* (blogspot.gr)

Me apáti éjinan **psefto-"ðiáloji"** me tis ámesa plitómenes kinonikés omáðes, ótan i vasikés apofásis ítan íði iliménes.

'**Pseudo-"dialogues"** with the directly affected social groups, when the basic decisions had already been taken, were made by fraud.'

To sum up, in (21)–(25) *pseudo(-)*X belongs to the category, but it is a negatively evaluated exemplar.

5.1.3 Undecidable (vague) reading

When associated with inherently vague predicates (Williamson 1975), that is predicates whose referential extension is vague, *pseudo(-)* signals that the referent is a borderline case of the category. Thus, the membership is logically undecidable. *Pseudo(-)* surfs on the vague, so the speaker's position is necessarily subjective and negatively orientated. *Pseudo(-)* does not solve the problem of the given interpretation; on the contrary, it focuses on the vague and more particularly on the devaluing aspect of *pseudo(-)*X.

The undecidable reading is mostly triggered when *pseudo(-)* combines with vague predicates (affects, qualities, properties, human status names), as illustrated in (26)–(30):

(26)   *Au tout début d'une vie amoureuse, nous rencontrons – et cela est difficilement avouable et acceptable – ce que j'appelle les **pseudo-amours**. Amours de besoin: je lui dis « Je t'aime » et cela veut surtout dire « Aime-moi ». Amours de peur: je lui crie « Je t'aime » et cela veut dire « Ne me quitte pas ».* (psychologies.com)

'At the very beginning of a love life, we encounter - and this is difficult to admit and to accept - what I call **pseudo-love**. Needy loves: I tell her "I love you" and that mostly means "Love me". Fear loves: I shout "I love you" and that means "Don't leave me".'

(27)   *Et des coups de force successifs vont fragiliser davantage ce peuple affamé par des **pseudos intellos** sans conscience des souffrances du peuple.* (france24.com)

'And successive coups de force will further weaken this starving people by **pseudo intellectuals** with no awareness of the suffering of the people.'

(28) *Vous avez un talent indéniable, et par conséquent vous méritez amplement d'être considéré comme un artiste, un vrai, pas comme ces **pseudos artistes** de pacotille qui vivent sur le travail des autres.* (simondaval.book.fr)

'You have an undeniable talent, and therefore you deserve to be considered an artist, a real one, not like these **pseudo artists** who live on the work of others.'

(29) *ήσαν καταραμένες, ως τάχα δαιμονικές ψευδαισθήσεις, ενός νέου "σύμπαντος" που απαιτούσε στο εξής την φυσική εξόντωση των "απίστων" στο όνομα του ψευτο-"Θεού" της **ψευτο-"Αγάπης"*** [...] (ysee.gr)

ísan kataraménes, os táxa ðemonikés psevðesθísis, enós néu "símbandos" pu apetúse sto eksís tin fisikí eksóndosi ton "apíston" sto ónoma tu psefto-"Θeú" tis **psefto-"Αγápis"** [...]

'(they) were cursed, as allegedly demonic illusions, of a new "universe" that henceforth demanded the physical extermination of "infidels" in the name of the pseudo-"God" of **pseudo-"Love"** [...]'

(30) *Ίσως όχι ο Δημητράς προσωπικά, αλλά έχει συμβεί αυτό με άλλους **(ψευτο) διανοούμενους** σε πάμπολλες περιπτώσεις σε όλες τις χώρες του κόσμου και σε όλα τα κράτη.* (ndimou.gr)

ísos óxi o Ðimitrás prosopiká, alá éxi simví aftó me álus **(psefto)ðianoúmenus** se pámboles periptósis se óles tis xóres tu kózmou ke se óla ta kráti.

'Maybe not Dimitras himself, but it has happened with other **(pseudo) intellectuals** on numerous occasions in all countries of the world and in all states.'

Examples (26)–(30) include occurrences, where the writers consider themselves incapable of drawing the boundaries of X[22], as it is difficult to say where the concept of 'love' (26), (29), 'intellectual' (27), (30) or 'artist' (28) begins and where it ends.

5.2 Ambiguous cases: Several categorizations, several readings

In this section we are concerned with nouns for which *pseudo(-)* – out of context – can target both category inclusion and category exclusion. The categories are clearly established, the opposable criteria can then be objective (*pseudo(-)*X is not an X), but they can also be subjective and play on connotations (*pseudo(-)*X is an X, but a bad example). In this case, the context makes it possible to know what type of categorization is involved (as demonstrated in (31)–(38) below and further discussed in what follows).

---

[22] Note that in (28) the writer explicitly points towards a no clear cut between a real and a not real (fake) artist and he uses the double hedging *real/pseudo(-).*

Thus, a third scenario may occur: in the absence of an enlightening context, *pseudo(-)*X can give rise to an ambiguous reading. Hence, there are potentially two types of categorization,

- *Pseudo(-)*X is an X.
- *Pseudo(-)*X is not an X.

and three types of possible readings:

- *Pseudo(-)*X does not belong to the category.
- *Pseudo(-)*X belongs to the category but is subjectively modalized.
- It is not known whether *pseudo(-)*X belongs to the category or not.

The trend[23] in our corpora is towards depreciative readings, i.e., subjectively modalized Xs. Human nouns and human collective nouns referring to institutions are the main types appearing in configurations raising several categorizations and several readings (see among others the French and Greek equivalents of *committee*, *team*, *association*, *organization*, *husband*, *tourist*, *supporter*, etc.) illustrated indicatively by 'team' in (31)–(32):

(31)   *Après la pitoyable défaite des Bleus face au Mexique, jeudi, Raymond Domenech ne trouvait pas de mots. La lecture de la presse du jour devrait grandement aider le sélectionneur à qualifier la bouillie de football offerte par sa **pseudo-équipe**. Le foutage de gueule est terminé, semblent dire les médias français. « Les imposteurs », titre* L'Equipe (maxifoot.fr)

'After the pitiful defeat of Les Bleus against Mexico on Thursday, Raymond Domenech was at a loss for words. Reading the press of the day should greatly help the coach to qualify the soccer mush offered by his **pseudo-team**. The French media seem to be saying that the bullshit is over. "The imposters", headlines L'Equipe'

(32)   *Il réalise son premier film, Sri Lanka National Handball Team, tourné en 2007 au Sri Lanka en langue cingalaise (et en Allemagne), d'après l'histoire vraie d'une **pseudo-équipe** de handball srilankaise qui s'est rendue en Bavière en septembre 2004 pour y disputer un tournoi avant de disparaître au bout de trois matches.* (wikipedia.org)

'He made his first film, Sri Lanka National Handball Team, shot in 2007 in Sri Lanka in the Sinhalese language (and in Germany), based on the true story of a Sri Lanka handball **pseudo-team** that went to Bavaria in September 2004 to play a tournament before disappearing after three matches.'

---

[23] A descriptive analysis of our data is beyond the scope of this paper due to space limitations; thus, this claim seems rather impressionist.

Pseudo(-)équipe 'pseudo-team' out of context can designate a real team but of poor quality, a fake team pretending to be a real team, or leave the interpretation undetermined. We are dealing here mostly with context-dependent occasional uses. 'Pseudo-team' in (31) is in fact the official French football team whose performance during the World Cup was criticized. In (32), it is obvious that the writer argues about a fake team "created" for the needs of a film production.

In the same vein, given the semantic properties of these nouns, there are cases where 'pseudo-tourist/husband/supporter' are effectively tourists, husbands, supporters: 'supporters' in (33) are counted in the ranks of the followers but in reality, they support Friedrich von Hayek by opportunism and do not behave like honorable supporters; in (34) the girl is officially married, but not to a man, i.e., not following the norms, or something like that:

(33) *Μαζί τους, αυτές τις συνέπειες, δεν θα τις αποφύγουν, ούτε οι σύγχρονοι Τσολάκογλου και Λογοθετόπουλοι, ούτε οι νεοφιλελεύθεροι **(ψευδο)οπαδοί** του Friedrich von Hayek, οι οποίοι κατήντησαν να συμπεριφέρονται ως ευκαιριακοί και εκ πεποιθήσεως, νεοκατοχικοί "κουκουλοφόροι" καταδότες της τρόικας των δανειστών.* (blogspot.gr)

Mazí tus, aftés tis sinépies, ðen θa tis apofíɣun, úte i síxroni Tsolákoɣlu ke Loɣoθetópuli, úte i neofileléfθeri **(psevðo)opaðí** tu Friedrich von Hayek, i opíi katíndisan na simberíféronde os efkeriakí ke ek pepiθíseos, neokatoxikí "kukulofóri" kataðótes tis tróïkas ton ðanistón.

'With them, these consequences will not be avoided, neither by the modern Tsolakoglou and Logothetopoulos, nor by the neoliberal **(pseudo)supporters** of Friedrich von Hayek, who have become accustomed to behave as opportunistic and by conviction, neophytes "hooded" informers of the lenders' troika.'

(34) *Il est même arrivé que la jeune fille fût assez innocente pour épouser une invertie, et pour vivre longtemps avec son **pseudo-mari** sans se douter qu'elle n'avait pas affaire à un homme.* (Beauvoir, *Le Deuxième sexe. II*, 1949)

'It even happened that the girl was innocent enough to marry an invert, and to live for a long time with her **pseudo husband** without suspecting that she was not dealing with a man.'

Similarly, we could cite a great number of instances where *pseudo(-)*X is not X, however, we restrict here to indicative examples as (35), where *pseudos touristes* 'pseudo tourists' pass off as tourists but are illegal workers, (36) where the man pretends to be the

husband, i.e., he plays a role and (37) where previous context indicates that it is a story about identity theft:

(35) *Ils découvrent que les **pseudos touristes** sont surtout venus pour couper des ananas pour cinquante euros par jour.* (clicanoo.com)

'They discover that the **pseudo-tourists** have mostly come to cut pineapples for fifty euros a day.'

(36) *Des journalistes new-yorkais se sont dit : « On va les avoir. On va prouver que ce n'est qu'un coup publicitaire. » Ils ont trouvé une femme enceinte et lui ont fait répéter un rôle, accompagnée de son **pseudo-mari**. Ce dernier est arrivé et a dit : « Ma femme est enceinte, mais elle veut voir Psychose. Laissez-nous entrer, le film a commencé s'il vous plaît. »* (wikipedia.org)

'Some New York reporters said, "We'll get them. We'll prove it's just a publicity stunt." They found a pregnant woman and had her rehearse a part, accompanied by her **pseudo-husband**. He came in and said, "My wife is pregnant, but she wants to see Psycho. Please let us in, the movie has started."'

(37) *Φοβούμενη πως ο **ψευτο-συγγενής** μπορεί να θέλει να τη βλάψει, η οικογένεια του Symansky περίμενε τρία χρόνια πριν τολμήσει τελικά να επικοινωνήσει με τις αρχές τον περασμένο Απρίλιο.* (flashnews.gr)

Fovúmeni pos o **psefto-sigenís** borí na θéli na ti vlápsi, i ikojénia tu Symansky perímene tría xrónia prin tolmísi teliká na epikinonísi me tis arxés ton perazméno Aprílio.

'Fearing that the **pseudo-relative** might want to harm her, Symansky's family waited three years before finally daring to contact authorities last April.'

Finally, notice that we cannot know whether we are talking about a real or a fake husband, tourist, etc. when the context is not informative enough as in (38)–(39):

(38) *Je m'inclinai seulement devant le **pseudo-mari** qui avait l'air fort contrarié, n'avait pas dit un mot, et trouvait cette conversation grotesque* (Havet, *Journal 1919-1924*, 2005)

'I only bowed to the **pseudo-husband** who looked very upset, did not say a word, and found this conversation grotesque'

(39) *Δεν έφταναν όλα τα άλλα, της βγήκε και ζηλιάρης ο **ψευτοσύζυγος**[24]*. (Ζ. Λαπνό, *Η Ελένη και το τέρας*, 2014)

Đen éftanan óla ta ála, tis vjíke ke ziliáris o **pseftosíziɣos**.

'As if everything wasn't enough, her **pseudo-husband** turned out to be jealous as well.'

---

[24] In Greek, to clearly mean *fake*, the unambiguous adjective *ψεύτικος* pséftikos (litt. 'false', 'fake') is used as in ***Ψεύτικο** ραντεβού με τον **ψεύτικο** σύζυγό σου;* (**Pséftiko** randevú me ton **pséftiko** síziɣó su? 'A **fake** appointment with your **fake** husband?'; OpenSubtitles2018.v3.). In (39), the speaker considers to be married to a 'pseudo husband' because of their inexistent love life.

As briefly demonstrated, all three readings of *pseudo(-)* are available with the nouns discussed in this section and, most of the times, the context helps us decide upon their specification.

6. Discussion

Pseudo(-), both in Greek and French, operates on the semantic as well as on the pragmatic dimension of utterances and a tendency towards devaluation is observed. It signals above all a "disproximation", a gap between *pseudo(-)*X (a comparator) and a compare X (Treis & Vanhove 2017) which, depending on the semantic and categorial potential of the noun, offers a more or less wide range of interpretative possibilities. From this point of view, we have argued that *pseudo(-)* is not an expression that makes a categorization fuzzy. It can clearly be used to assert a categorial exclusion on the basis of a semantic mismatch, it can be used to negatively evaluate a member of a category, and it can finally signal an exemplar that is considered to be a borderline case of a vague category.

Even if semantic and pragmatic dimensions are hard to tease apart, what seems to prevail is that *pseudo(-)* is a vague term itself, in the sense of semantically underspecified, making in many contexts the interpretation of *pseudo(-)*X undecidable or oscillating between devalued belonging and not belonging. As such, one cannot decide in advance whether the vagueness is "intentional" or not (Voghera 2012). The choice of the X noun in the right context of *pseudo(-)* proves to be relevant as it can specify *pseudo(-)*'s semantic orientation. Thus, the semantic features of the noun allow us to build on a predictive model towards untangling propositional content and illocutionary force of *pseudo(-)*X. So, although *pseudo(-)* is a hedge, it is not forced to inject vagueness into the *pseudo(-)*X sequence. It is not even forced to introduce subjectivity into the discourse. It is, however, sensitive to the element X on which it focuses. The model we presented is not intended to be exhaustive, but rather to show that it is possible, on the basis of the semantic and categorial properties of elements X, to understand when the scope of *pseudo(-)* is semantic (see for instance the '*pseudo ice floe*' on which we find a bear in a zoo that is objectively *not an ice floe* or a '*pseudo husband*' who can designate someone who is not objectively the

husband) and when it is pragmatic (see a '*pseudo husband*' who can designate a husband but subjectively depreciated), or both.

Still, much work remains to be done. We need to analyze the interesting double marking on *pseudo(-)* (i.e., *pseudo(-)* in quotation marks or parenthesis: "pseudo"-scientist or (pseudo)-scientist) that could reveal a form of linguistic insecurity towards its status, as shown in Section 2, along with an indication of a subjective designation. More research must be also conducted by examining in a systematic and quantitative way the pairing between *pseudo(-)* and X examples appearing in the second position of the structure in order to reveal fine-grained affinities, recurrences and preferences of *pseudo(-)* and to provide statistics about the distribution of the readings. For instance, our corpus data do not include general nouns like *thing*, *action*, *movement*, *noise*, *smell*, *gesture*, etc. in the X position. If we considered examples like 'pseudo-noise' or 'pseudo-odor', the following questions would arise: what can sound like a noise without being a noise? Can we devalue the quality of a noise, of a smell, etc.? At this point, we could not think of something that would justify the presence of *pseudo(-)*.

Besides, one way of studying *pseudo(-)* and conceiving vagueness as the result of the (im)possibility to distinguish between what is true or false is to compare it with expressions which, by different ways, convey similar effects, that is blurring the boundaries of another linguistic expression when this is permitted by the noun. For example, in Greek, *miso(-)* 'semi', *koutso(-)* 'limping/lame', *psilo(-)* 'thin/a bit' can be in competition with *psefto(-)*/*psevðo(-)* (see Tsamadou-Jacoberger 2009; Xydopoulos 2009). In French (Gerhard-Krait et al. 2023), *simili(-)* is used in a quasi-complementary distribution with *pseudo(-)* (see also Masini & Micheli 2020 for Italian).

Specifically for Greek, we have to study in depth *pseudo(-)* associated to verbs and to examine some interesting cases where *pseudo(-)* is not fully grammaticalized, like *psefto-ponirákias* 'pseudo-cunning' mentioned in Section 2.2. We suggested that a possible explanation would be that *pseudo(-)* has been reanalyzed as an evaluative (depreciative) adjective (in contrast to what appears as a privative prefix).

Finally, we would like to stress that our study contributed to disentangle, among all that can generate vagueness in language, what is semantic vagueness (intrinsic) from what is pragmatic vagueness (the set of subjective modalities emanating from observational

vagueness, and from qualitative judgment). Thus, the semantic dimension (i.e., the speaker's intention plays no role here) must be taken into account along with the pragmatic one (i.e., judgment, subjectivity). The patent confusion between pragmatic vagueness and semantic vagueness led to the overpowering character of the prototype theory. Instead, we insisted in this paper on the necessity of considering the rigid vs. extensible nature of noun categories and thus the possibility of the categorial membership or inclusion of *pseudo(-)*X in the category X by verifying the capacity of X to constitute a category that can be grasped by approximation or not.

References

Amiot, Dany & Dejan Stosic. 2015. Morphologie évaluative et aspectuelle en français et en serbe. *Lexique* 22. 111–142.

Amiot, Dany & Dejan Stosic. 2022. Evaluative Morphology: From evaluation to approximation and semi-categorization. In Hélène Vassiliadou & Marie Lammert (eds.), *A Crosslinguistic Perspective on Clear and Approximate Categorization*, 53–94. Newcastle upon Tyne: Cambridge Scholars Publishing.

Anastassiadi-Symeonidi, Anna & Asimakis Fliatouras. 2019. Το λόγιο επίπεδο της σύγχρονης Νέας Ελληνικής: Συγχρονικές και διαχρονικές τάσεις [The literary register in Modern Greek: Diachronic and synchronic perspectives]. In Asimakis Fliatouras & Anna Anastassiadi-Symeonidi (eds.), *Το λόγιο επίπεδο στη σύγχρονη Νέα Ελληνική* [The literary register in Modern Greek], 15–48. Athens: Patakis.

Babiniotis, Georges. 1969. *Ο δια συνθέσεως υποκορισμός εις την ελληνικήν.* [Diminution in Greek compounding]. Ph.D. dissertation. University of Athens.

Baeskow, Heike. 2004. *Lexical Properties of Selected Non-native Morphemes of English.* Tübingen: Narr.

Cappelle, Bert, Robert Daugs & Stefan Hartmann. 2023. The English privative prefixes *near-*, *pseudo-* and *quasi-*: Approximation and 'disproximation'. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1). 52–75.

Corbin, Danielle. 1987. *Morphologie dérivationnelle et structuration du lexique.* Tübingen: Niemeyer.

Devos, Filip. 2003. Semantic vagueness and lexical polyvalence. *Studia Linguistica* 57(3). 121–141.

Efthymiou, Angeliki. 2017a. Η έκφραση της επίτασης και του μετριασμού στα αξιολογικά ρήματα της Νέας Ελληνικής. [The expression of intensity and attenuation within evaluative verbs in Modern Greek]. *Studies in Greek Linguistics* 37. 227–238.

Efthymiou, Angeliki. 2017b. Intensification and deintensification in Modern Greek verbs. *Lexis-Journal in English Lexicology* 10. DOI: 10.4000/lexis.1089.

Efthymiou, Angeliki. 2019. Οι σημασίες των μετριαστικών ρημάτων στη Νέα Ελληνική. Ένα σύνθετο παζλ. [Meanings of diminutive Greek verbs. A complex puzzle]. In Grammatikí Karla, Io Manolessou & Nikolaos Pentelidis (eds.), *Lexeis. In honour of Christina Basea-Bezentakou*, 69–84. Athens: Kardamitsa.

Flaux, Nelly & Danièle Van de Velde. 2000. *Les noms en français*. Paris: Ophrys.

Flaux, Nelly & Dejan Stosic. 2015. Pour une classe des noms d'idéalité. *Langue française* 185. 43–57.

Fotiadou, Georgia, Francine Gerhard-Krait, Marie Lammert & Hélène Vassiliadou. Forthcoming. *Pseudo(-)* in Greek: A morpheme in categorization, approximation and depreciation. *Studies in Greek Linguistics* 42.

Gerhard-Krait, Francine & Hélène Vassiliadou. 2014. Lectures taxinomique et/ou floue appliquées aux noms: quelques réflexions... *Travaux de linguistique* 69. 57–75.

Gerhard-Krait, Francine & Hélène Vassiliadou. 2017. *Clapotis*, *murmures* et autres manifestations sonores: les méandres de l'approximation catégorielle. *Syntaxe et Sémantique* 18. 19–43.

Gerhard-Krait, Francine & Maria Zerva. 2023. La catégorisation et ses fluctuations formelles et interprétatives. Introduction. *Langages* 229. 9–30.

Gerhard-Krait, Francine, Marie Lammert, Hélène Vassiliadou & Maria Zerva. 2023. Catégorisation et approximation à la manière de *simili(-)*X. *Langages* 229. 37–56.

Giannoulopoulou, Giannoula. 1997. Συμφύματα στη Νέα Ελληνική. Η περίπτωση του *ψεύδο-*. [Confixes in Modern Greek: the case of *pseudo-*]. *Studies in Greek Linguistics* 17. 115–129.

Haspelmath, Martin. 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language* 86(3). 663–687.

Huyghe, Richard. 2015. Les typologies nominales: présentation. *Langue française* 185. 5–27.

Kahlaoui, Karima, Bernadette Ska, Clotilde Degroot & Yves Joanette. 2010. Neurobiological bases of the semantic processing of words. In Jackie Guendouzi, Filip Loncke & Mandy J. Williams (eds.), *The Handbook of Psycholinguistic and Cognitive Processes*, Chapter 5. Abington: Routledge.

Kleiber, Georges. 1987. Quelques réflexions sur le vague dans les langues naturelles. In Sylvie Mellet (ed.), *Etudes de linguistique générale et de linguistique latine offertes en hommage à Guy Serbat*, 157–172. Paris: Bibliothèque de l'Information grammaticale.

Kleiber, Georges. 2001. Remarques sur la dénomination. *Cahiers de praxématique* 36. 21–41.

Körtvélyessy, Lívia. 2015. *Evaluative morphology from cross-linguistic perspective*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Lakoff, George. 1973. Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic* 2(4). 458–508.

Lieber, Rochelle. 2004. *Morphology and Lexical Semantics*. Cambridge: University Press.

Masini, Francesca & M. Silvia Micheli. 2020. The morphological expression of approximation: The emerging *simili-* construction in Italian. *Word Structure* 13(3). 371–402.

Melissaropoulou, Dimitra. 2015. Modern Greek. In Nicola Grandi & Lívia Körtvélyessy (eds.), *Edinburgh Handbook of Evaluative Morphology*, 269–277. Edinburgh: University Press.

Melissaropoulou, Dimitra & Angela Ralli. 2020. Derivational networks in Greek. In Lívia Körtvélyessy, Alexandra Bagasheva and Pavol Štekauer (eds.), *Derivational Networks Across Languages*, 347–370. Berlin: De Gruyter Mouton. DOI: 10.1515/9783110686630-035.

Mihatsch, Wiltrud. 2009. L'approximation entre sens et signification: un tour d'horizon. In Dominique Verbeken (ed.), *Entre sens et signification - Constitution du sens: points de vue sur l'articulation sémantique-pragmatique*, 100–116. Paris: L'Harmattan.

Mihatsch, Wiltrud, Inga Hennecke, Anna Kisiel, Alena Kolyaseva, Kristin Davidse & Lieselotte Brems (eds.). 2023. *Type Noun Constructions in Slavic, Germanic and Romance Languages. Semantics and Pragmatics on the Move*. Berlin: Mouton de Gruyter. Trends in Linguistics 352.

Norde, Muriel. 2009. *Degrammaticalization*. Oxford: Oxford University Press.

Quirk, Randolf, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

Raffman, Diana. 2011. Vagueness and Observationality. In Giuseppina Ronzitti (ed.), *Vagueness: A Guide*, 107–121. New York: Springer.

Ralli, Angela. 2004. Stem-based versus word-based morphological configurations: The case of Modern Greek preverbs. *Lingue e Linguaggio* 2. 241–275.

Ralli, Angela. 2005. *Μορφολογία*. [Morphology]. Athens: Patakis.

Scalise, Sergio. 1990. *Morfologia e lessico. Una prospettiva generativista*. Bologna: Il Mulino.

Schmid, Hans-Jörg. 2000. *English Abstract Nouns as Conceptual Shells: From Corpus to Cognition*. Berlin & New York: Walter de Gruyter.

Štekauer, Pavol, Salvador Valera & Lívia Körtvélyessy. 2012. *Word-Formation in the World's languages. A Typological Survey*. Cambridge: Cambridge University Press.

Stosic, Dejan & Nelly Flaux. 2021. Les noms d'idéalité et les noms sous-spécifiés. *Corela* HS 34, DOI: 10.4000/corela.13587 (accessed January 2022).

Thillmann, Sandra. 2007. A linguistic analysis of the word element 'pseudo', *Seminar Paper*, Munich, GRIN Verlag. https://www.grin.com/document/88693 (accessed December 2021).

Treis, Yvonne & Martine Vanhove (eds.). 2017. *Similative and Equative Constructions: A Cross-linguistic Perspective*. Amsterdam: John Benjamins.

Tsamadou-Jacoberger, Irini. 2009. L'élément *μισο-* et la notion de frontière en grec moderne. In Zoé Gavrielidès (ed.), *Actes du XXe Colloque International des Néo-hellénistes des Universités francophones*, 579–588. Lille: Septentrion.

Van Goethem, Kristel & Muriel Norde. 2020. Extravagant "fake" morphemes in Dutch. Morphological productivity, semantic profiles and categorical flexibility. *Corpus Linguistics and Linguistic Theory* 16(3). 425–458. DOI: 10.1515/cllt-2020-0024 (accessed January 2022).

Van Goethem, Kristel, Muriel Norde & Francesca Masini. 2021/under review. The fate of 'pseudo-' words: a contrastive corpus-based analysis. Paper presented at UCCTS 2021 (Using Corpora in Contrastive and Translation Studies), Bertinoro, 9–11 September 2021.

Vassiliadou, Hélène & Marie Lammert (eds.). 2022. *A Crosslinguistic Perspective on Clear and Approximate Categorization*. Newcastle upon Tyne: Cambridge Scholars Publishing.

Vassiliadou, Hélène, Elena Vladimirska, Marie Lammert, Céline Benninger, Francine Gerhard-Krait, Jelena Gridina & Daina Turla. 2023. Clear vs. approximate categorization in French and Latvian. In Wiltrud Mihatsch, Inga Hennecke, Anna Kisiel, Alena Kolyaseva et al. (eds.), *Type Noun Constructions in Slavic, Germanic and Romance Languages. Semantics and Pragmatics on the Move*. Chapter 17. Berlin: Mouton de Gruyter. Trends in Linguistics 352.

Voghera, Miriam. 2012. Ghitarre, violini, banjo e cose del genere. In Anna Thornton & Miriam Voghera (eds.), *Per Tullio De Mauro: studi offerti dalle allieve in occasione del suo 80 compleanno*, 341–364. Roma: Aracne.

Williamson, Timothy. 1975. On the coherence of vague predicates. *Synthèse* 30. 325–363.

Xydopoulos, George. 2009. *Psilo-*: Exploring the case of a 'diminutive' in Modern Greek. In Anastasios Tsangalidis (ed.), *Selected papers from the 17th International Symposium of Theoretical and Applied Linguistics*, 397–405. Thessaloniki: Monochromia.

Corpora

Frantext, ATILF: www.frantext.fr (accessed October 2021).

French Web Corpus 2017, frTenTen17: app.sketchengine.eu (accessed October 2021).

Greek Web Corpus 2014, elTenTen14: app.sketchengine.eu (accessed November 2021).

Hellenic National Corpus (*ILSP*): corpus.ilsp.gr/index.php (accessed November 2021).

Leipzig Corpora Collection – Wortschatz, corpora.uni-leipzig.de (accessed November 2021).

Thesaurus Linguae Graecae© Digital Library. Ed. Maria C. Pantelia. University of California, Irvine. stephanus.tlg.uci.edu (accessed February 2022).

Hélène Vassiliadou

University of Strasbourg / UR 1339 Linguistique, Langues, Parole

Faculté des Lettres

14 rue René Descartes

BP 80010

67084 Strasbourg cedex, France

vassili@unistra.fr

Francine Gerhard-Krait

University of Strasbourg / UR 1339 Linguistique, Langues, Parole

Faculté des Lettres

14 rue René Descartes

BP 80010

67084 Strasbourg cedex, France

gerhard@unistra.fr

Georgia Fotiadou

Aristotle University of Thessaloniki

School of Philology, Department of linguistics

Old building of Philosophical School, room 301

54124 Thessaloniki, Greece

geofotia@lit.auth.gr

Marie Lammert

University of Strasbourg / UR 1339 Linguistique, Langues, Parole

Faculté des Lettres

14 rue René Descartes

BP 80010

67084 Strasbourg cedex, France

mlammert@unistra.fr

Miriam Voghera

The role of diminutive suffixes in the Italian Time Nouns constructions: From approximation to focus?

Abstract: The time nouns (TNs), *minuto* ('minute'), *momento* ('moment'), *attimo* ('instant') can be part of constructions in which they keep all their categorial and syntactic properties, but also in constructions which seem to undergo a process of grammaticalisation: loss of categorial property of nominal inflection; the sequence DET+TN (*un attimo, un momento, un minuto*) can no longer be interrupted by modifiers nor have postnominal modifiers; semantic shift from the original temporal meaning to the more general meaning of a '(small) indefinite amount', to continue towards more properly grammatical values of quantifiers to pragmatic values of hedges, alerter and textual marker of focus. Interestingly, while the original temporal constructions do not occur in diminutive forms, the departure from temporal meaning makes the use of diminutive forms possible. The article explores in which cases they are allowed and the semantic and textual factors which contribute to their presence.

Keywords: evaluative morphology, diminutives, intentional vagueness, focalising constructions, time noun constructions, Italian

1. Introduction[*]

This article deals with some Italian constructions with time nouns (TNs), *minuto* ('minute'), *momento* ('moment'), *attimo* ('instant') that, besides expressing a temporal meaning, have developed functions and meanings ranging from intentional vagueness to focus, as it happens to many pragmatic markers (Underhill 1988; Miller & Weinert 1995; Mihatsch 2020; Voghera 2017, 2022, 2023; Voghera & Borges 2017). In particular, I investigate why the diminutive forms of the considered TNs are only allowed in some TN constructions (CxsTN).

The considered TNs can enter into constructions that play the role of temporal adjuncts either as NP or as prepositional complements within a PP (CxsTN[TEMP]). In this case they maintain their categorial features and original meaning, as in the following examples:

(1)   *Tutto è successo in un momento/attimo/minuto*
        'It all happened in a moment/instant/minute'

(2)   *Ti aspetto da sette minuti/*attimi/*momenti*
        'I've been waiting for you for seven minutes/instants/moments'

(3)   *Un attimo/momento/*minuto e il ladro è scappato*
        'One instant/moment/minute and the thief ran away'

(4)   *Tutto è successo in un momento/*attimo/*minuto complicato*
        'It all happened during a complicated moment/instant/minute'

As we can see, only in (1) the TNs are interchangeable. In (2), where the TN is preceded by the numeral *sette* ('seven'), only *minuto* ('minute') can occur because it is a "noms du référentiel chronologique" (Berthonneau 1989: 399), i.e., a measurable unit of time (Fillmore 2002), the meaning of which is gradable and segmentable in discrete units: saying *seven minutes* we mean a defined portion of time. On the contrary, *attimo* ('instant') and *momento* ('moment'), unlike *minuto*, have no definite relation to a specific referential class of temporal units; their meaning is non-gradable, and it is not possible to indicate a part of *attimo* or *momento*: one cannot say half an *attimo* or half a *momento*. This the reason why in (3) – where the speaker wants to express instantaneity as a whole – only *attimo* and *momento* are acceptable. Finally, in (4) only *momento* is acceptable because the presence of the adjective *complicato* ('complicated') leads to a durative interpretation and *momento*, differently from *attimo* and *minute*, can also mean 'period'.

Now let's see what happens if we use the same sentences with the TNs in the diminutive form. To test the acceptability of diminutive forms in the different constructions, I checked all the contexts in which the diminutives of the TNs occur in the two main corpora of spoken Italian, VoLIP and Kiparla, and the first thousand contexts in the corpus of digital

written Italian ItTenTen 2020[1], which represent colloquial and informal register of the language, albeit written.

(5)   *Ti aspetto da sette minutini*
'I've been waiting for you for seven minutes.DIM'

(6)   *Un attimino/*momentino e il ladro è scappato*
'One instant.DIM/moment.DIM and the thief ran away'

(7)   *Tutto è successo in un momentino*
'It all happened in one moment.DIM'

(8)   *Tutto è successo in un momentino complicato*
'It all happened during a complicated moment.DIM'

All the sentences result unacceptable, but (8), although I did not find any example of this type in the consulted corpora. Their unacceptability does not depend on the fact that TNs cannot be used in the diminutive form because indeed, as we shall see, diminutive forms are quite frequent. They occur, however, in constructions that no longer have their original temporal meaning, but acquire that of expressions of intentional vagueness or focusers. This difference in the use of diminutives in different types of constructions is, I believe, interesting in order to better understand their meaning and function.

In Section 2, I will briefly present the meanings and functions of diminutives in Italian and in particular of the TNs under analysis; in Section 3, I will analyse the meaning and the functions of TNs constructions that admit diminutives; in Section 4, finally, I will draw some conclusive remarks.

2. A look at diminutives

As is well known, evaluative morphology in Italian is very rich. To get an idea of the variety and number of suffixes that are used for diminutives, I report the list in Merlini Barbaresi (2004: 265–266):

---

[1] The meaning of diminutive forms is subject to synchronic variation (see Section 2) and in Italian the acceptability of utterances is also strongly dependent on diatopic factors: in some dialects and regional varieties of Italian they are more frequent and therefore more acceptable. Thus, I am aware that the evaluation I present may have partially different degrees of acceptability in speakers with other regional backgrounds.

-ino/a, -etto/a, -ello/a, -uccio/a, -uzzo/a, -otto/a, -(u)olo/a, -icci-, (u)olo/a, -iolo/a, -acci-olo, -olo/a, -àtt-olo/a, -onz-olo/a, -usc-olo, -agn-olo, -ign-olo/a, -occ-olo, -isc-olo, -ùgi-olo/a, -icola/o, -occhio/a, - occio/a, -ozzo/a, -atto/a, -acchio/a, -icchio/a, -ulo/a, -iggine, -iglio, -ecchio, -ischio, -ottero/a

Most of the literature on diminutive suffixes agrees on their polysemy (Jurafsky 1996; Delhay 1999; Grandi & Körtvelyéssy 2015; Prieto 2015) and on the fact that they can have scope over different domains (Dressler & Merlini Barbaresi 1994). Grandi and Körtvelyéssy (2015: 10) distinguish between two perspectives: descriptive and qualitative. The former "relies on real and objective properties of an item", the latter "assumes a subjective evaluation: personal feelings or opinions and, often, the influence of extra-linguistic context becomes the crucial factor". This distinction is intuitively very convincing, but not always easy to apply. In fact, any non-metrical evaluation implies, in my opinion, always subjectivity, after all the semantic primitives, 'small' and 'big', to which we refer to identify the basic meaning of diminutives or augmentatives are vague predicates and therefore their meaning is always valid only in respect to a reference parameter.

The proposal in Grandi (2017) seems more effective, where it is stated that the basic value for all evaluative affixes is that of 'deviation', in various possible directions, from the meaning expressed by the base form. According to Grandi, diminutives approximate by defect, augmentative by excess.[2] In this perspective, 'small' for diminutives and 'big' for augmentatives are only two of the possible dimensions of approximation and not the primary ones. This proposal, also supported by diachronic data (Grandi 2017: 148–151), eliminates the primacy of the dimensional meaning and has the advantage of being applicable also to meanings that are not inherently gradable.

The most interesting point, from my perspective, is that Grandi's proposal can eliminate the distinction between 'objective' and 'subjective' deviation. In fact, it seems more convincing to claim that the different meanings expressed by diminutives are always expressions of speaker's attitude but can be arranged along a continuum that goes from a shared intersubjectivity to a highly subjective meaning. In considering the value of diminutives, therefore, two elements intersect: on the one hand, the semantics of the base that certainly

---

[2] For a different opinion about augmentatives see Stosic & Amiot (2023).

pushes the interpretation of diminutives towards some specific dimension of approximation and, on the other, the degree of shared subjectivity.

Intersubjectively
shared meaning

$\longleftrightarrow$

Highly subjective
meaning

fazzolettino
'handkerchiefDIM/small'

maritino
'husbandDIM/ hubby'

Fig. 1: Continuum of deviation values expressed by diminutives

If Italian speakers use the word *fazzolettino* ('handkerchief.DIM'), we can reasonably assume that they have a shared representation 'small handkerchief', but if we use the word *maritino* ('hubby', lit. husband.DIM), it is much more difficult to establish a shared representation, if there is a shared representation at all, also because in this case the diminutive can have both positive and ironic connotations.

The continuum between shared intersubjectivity and high subjectivity avoids having to identify a unique feature for all non-lexicalised diminutives, such as [non-serious] proposed by Dressler and Merlini Barbaresi (1994), which, according to the authors, is a combination of [fictive] and [non-important]. In fact, it seems that this feature would not apply to diminutives such as *fazzolettino*, *pezzetto* ('piece.DIM') and many others.

Another point in favour of Grandi's (2017) proposal is the consideration of diminutives as approximate meanings of the base, which emphasises the comparison that underlies the meanings expressed by evaluative suffixes (Grandi & Körtvelyéssy 2015: 13; Tovena 2015: 118). Even more explicit is the position of Delhay (1999: 80):

> En effet, on peut considérer que l'opération dérivationnelle "diminutive" met en jeu un couple X / Xd et que le formant d, généralement un suffixe "diminutif", pose l'existence d'un écart sémantique entre la dénomination X et la dénomination Xd, écart que j'ai propose d'appréhender par la paraphrase "Xd n'est pas un vrai X". On constate que les dénominations Xd [...]comportent toutes un composant sémantique d'ordre relationnel (sorte de, espèce de, analogue à, qui ressemble à, en relation avec, fait de, partie de, etc.).

This interpretation of diminutives' meaning is of particular interest because in various studies it has been shown that more or less explicit comparative processes are the basis for the development of many vagueness expressions (Voghera 2014, 2022, 2023).

The TNs taken into consideration here all admit the diminutive with the suffix *-ino*, which is also the most frequent and productive one in Italian. As we have already seen, *-ino* approximates the meaning of the base in different ways not necessarily related to the concept of smallness (*fazzolettino* vs. *maritino*). However, it has an inherent meaning of smallness/diminutive, which is made evident in cases where *-ino* can be used as a free-standing word to intensify also diminutives with other suffixes:

(9)   A: *Vuoi della torta?*
                  'Do you want some cake?'

          B: *Sì, un **pezz+etto**, ma **ino***
                  'A little piece.DIM, but really little'

In this case the use of *ino* means 'really small' and has the function of intensifying the meaning 'small' already expressed by the suffix *-etto* in *pezzetto*.

The possibility of using diminutives with bases denoting metric measures, *minutino* ('minute.DIM'), *chiletto* ('kilo.DIM'), *settimanella* ('week.DIM') etc., is from a certain point of view, a paradox. In such cases it is difficult to find a single meaning for the diminutive forms and the context is a decisive key, as we shall see. What can be said, however, is that neither *minutino* nor *settimanella* nor *chiletto* necessarily indicate an approximation by defect, since they can be interpreted as '±base's meaning'. In these cases, it is questionable whether they can be always considered hyponyms of the base.

3. From CxsTN[TEMP] to intentional vagueness

The three TNs can enter in constructions that express intentional vagueness. We have intentional vagueness when the speaker more or less consciously makes linguistic choices with low discriminating power in relation to the situation:

> un'espressione è vaga quando non possiamo decidere in base a considerazioni formali se, noto il referente e nota l'espressione, essa è applicabile sempre o non è applicabile mai al referente (De Mauro 1982: 99)[3]

---

[3] Translation MV: 'an expression is vague when we cannot decide on the basis of formal considerations whether, given the referent and the expression, it is always applicable or never applicable to the referent'.

Intentional vagueness consists of constructions that do not necessarily depend on systemic factors (Voghera 2012, 2013, Voghera & Collu 2017), as in the following example:

(10)   You know, John is bald, something like that

Bald is a vague predicate, the vagueness of which depends on the linguistic code, but the speaker adds more vagueness to the utterance with the Vagueness Expressions (VEs) *you know*, *something like that.*

Intentional vagueness responds to different communicative needs, on the basis of which I proposed to distinguish three kinds of vagueness (Voghera 2017, 2022; Voghera & Collu 2017):[4]

a.   informational vagueness, the domain of which is the propositional content, depending on lack of information:

(11)   *la terrazza sarà_ tre o quattrocento metri quadri* (VoLIP)
          'the terrace must be three or four hundred square meters'

b.   relational vagueness, the domain of which is the pragmatic relation between speakers and the propositional content, depending on difficulty or reluctance to establish a clear relation with the content of the utterance or the addressee:

(12)   *siamo un po' imbarazzati* (VoLIP)
          'we are a bit embarrassed'

c.   discourse vagueness, the domain of which is the textual construction, depending on exigencies due to online programming and production processes both in speech and in not prepared writing, such as notes:

(13)   A: oh you mean I know like I know every, I know. so many people now this
          year (Jucker et al. 2003: 1754)

The approach presented here has a lot in common with Caffi's view (2007) because of the holistic view of this kind of phenomena and because of the distinction between different kinds of vagueness (*mitigation* in Caffi's terms). Intentional vagueness, however,

---

[4] As pointed out by Masini, Norde and Van Goethem (2023) in the introductory section, many terms have been used to cover these kinds of phenomena. One of the most popular is vague language (Channel 2004; Cutting ed. 2007), which I find misleading because it leads one to believe that the terms or constructions used in these cases are vague per se, whereas on the contrary many VEs become such only in the specific context.

differs from mitigation in that it concerns both the expression of content and of form, manifested for example through hypoarticulation (Linbloom 1990), which I will not discuss here. Moreover, the use of the term mitigation seems to inevitably imply downgrading of some kind which is not always present in VEs.

The three types of intentional vagueness can combine and overlap in several contexts and a VE can convey different types of vagueness: for instance, *diciamo* in Italian or *sagen wir mal* in German ('let's say') can be used to convey all three kinds of vagueness (Caffi 2007; Mihatsch 2010; Voghera & Collu 2017).

It is not possible to reduce intentional vagueness to a limited set of linguistic-cognitive operations, but from the studies carried out in recent years (Channel 1994; Jucker et al. 2003; Mihatsch 2010, 2016, 2020; Overstreet 1999, 2011), three operations, with large areas of overlapping, seem to be highly recurrent: generalisation, approximation, attenuation.

(14)	A*: Cosa vuoi che ti porto stasera?* B*: Gin, robe del genere...*	Generalisation
	A: What shall I bring tonight? B: Gin, stuff like that...	
(15)	A: *Quante persone c'erano?* B: *Tre quattro*	Approximation
	A: How many people were there? B: Three four	
(16)	A: *A che ora vengo a cena?* B: *Mah, otto? nove?*	Attenuation
	A: What time shall I come for dinner? B: Well, eight? nine?	

All B utterances in the preceding dialogues have a low discriminating power due to the use of the implicit analogy expressed by the general extender *robe del genere* 'stuff like that' in (14), the unresolved alternative expressed by the construction with the number pair *tre quattro* 'three four' in (15) (Voghera 2019) and by the answer given in interrogative form in (16), which is a courtesy attenuative strategy because the speaker, instead of giving a straight answer, apparently leaves the decision on the time of dinner to the recipient.

Most VEs are polyfunctional. As noted in studies based on Italian, English, French, Spanish, Portuguese and German corpora, from a diachronic point of view, there is a path

of development, according to which expressions that originally express informational vagueness can evolve towards the expression of relational vagueness and subsequently towards the expression of discursive vagueness (Jucker et al. 2003; Mihatsch 2020; Ghezzi 2013; Voghera 2014, 2022; Voghera & Borges 2017; Voghera & Collu 2017). In other words, the expression of relational and discursive vagueness is expressed by refunctionalising constructions originally used to vehiculate information vagueness. Thus, the well documented diachronic path from the expression of propositional meanings to pragmatic-discursive meanings, works for the VEs as well (Traugott 1995, 1982, 2003; Davidse, Vandelanotte & Cuyckens (eds.) 2010; Diewald 2011; Ghezzi & Molinelli (eds.) 2014). The CxsTN$_{[TEMP]}$ seem to follow the same path, although a diachronic investigation is required to confirm it.

3.1 The approximative constructions: CxsTN$_{[APPROX]}$ and CxsTN$_{[QUANT]}$

All three TNs can be part of two different but strictly related approximative constructions, which admit the diminutive form in most cases.

CxsTN$_{[APPROX]}$ function as temporal approximators, in which the temporal meaning is completely indeterminate both from a semantic and an aspectual point of view, and their meaning can be paraphrased as '(small) indeterminate portion of time'. *Momento* and *attimo* occur only in the singular form with the indefinite article (*un attimo*, *un momento*), while *minuto* can occur in the plural form accompanied by small numerals which are used with an undetermined value, such as one, two, five (Voghera 2019). It is very frequent that they occur followed by PP specifiers, as in (18) and (20).[5]

(17)   *ecco io la blocco un attimo/momento/minuto perché mi dicono che c'è la pubblicità* (VoLIP)

          'here I'm stopping you for an instant/moment/minute/because they tell me there is advertising'

(18)   *calmo e fermo, solo un po' arrossato, mai un momento/attimo/minuto* [*di smarrimento o esitazione*]$_{PP}$ (VoLIP)

          'calm and steady, just a little flushed, never a moment's/instant's/minute's disorientation or hesitation'

---

[5] In all examples I put firstly the TN found in the corpus followed by the two other TNs marked by their degree of acceptability.

In these examples the CxsTN can reasonably mean '(small) indeterminate portion of time', but through a metonymic interpretation also '(small) indeterminate amount/a bit'. This second interpretation becomes the most probable if we change the semantic frame in which the construction is needed:

(19)   *vieni Nino mettiti un attimo/momento/minuto serio* (VoLIP)

'come Nino be a little bit serious' (lit. stay an instant/moment/minute serious)

(20)   *scusa c'è stato un momento/attimo/?minuto* [*di di di intrecci di fili del telefono*]ₚₚ (VoLIP)

'sorry there was a bit of intertwining of telephone wires (lit. sorry there was a moment/instant/minute of intertwining of telephone wires)

(21)   *quindi mi andrebbe di fare qualcosa un attimo/ *momento/ *minuto di attivo* (KIParla)

'so I'd like to do something a bit active (lit. something an instant of activity)'

Examples (19)–(21) show uses of different TNs, which go from temporal to quantity approximation. If we use *un minuto*, the temporal meaning always remains active and its insertion both in (19) and (20) would lead to interpret the construction as temporal approximation; where temporal interpretation is excluded, as in (21), the construction with *un minuto* is completely unacceptable.[6] The constructions with *un momento* are ambiguous because they can be both interpreted as meaning temporal approximation in (19) and quantitative approximation in (20). However, in (21), where TNcxs can only mean quantitative approximation, *un momento* is unacceptable. *Un attimo* is acceptable in all examples with a still possible temporal interpretation in (19) and a clear quantitative interpretation in (20)–(21). This means that *un minuto* has not shifted from temporal to quantitative approximation, *un momento* can only be used in contexts where both temporal and quantitative approximation are possible, while in the sentences with *un attimo* the distinction between the two kinds of approximation is blurred.

These constructions of quantificational approximation (CxsTN[APPROX]) triggered a path similar to the one already studied for other complex and degree modifiers originating from the NP of NP, on which there is an abundant bibliography (Traugott 2008; Masini 2012,

---

[6] As native speaker I could accept *minuto*.DIM also as quantitative approximator in some case, but no such example is present in the consulted corpora.

2016; Mihatsch 2010, 2016; Giacalone Ramat 2015; Voghera 2013). Italian examples are: *un sacco di* ('a sack of'), *una marea di* ('a tide of') and several others, which nowadays mean 'a lot'. In these kinds of constructions formed by NP of NP the meaning of the head of the first NP is bleached and loses all the categorial features of a noun. The path of reanalysis that thus allows the transition of the initial NP to quantifier is illustrated in Figure 2.

NP of NP	CxsTN[QUANT]	Examples
[*un attimo/momento*] NP *di* [NP]	[[*un attimo/momento di*]QUANT]NP	*un attimo/momento di intrecci di fili*
'an instant/moment of [NP]'	'a little bit, a bit of NP'	'a bit of intertwining of telephone wires '
	[[*un attimo/momento*]QUANTA]A	*un attimo/momento serio*
	'a little bit A'	'a little bit serious'

Fig. 2: Process of re-analysis of the sequence NP of NP

The diminutive is always acceptable in both Cxs[APPROX] and Cxs[QUANT] and its use affects neither their propositional meaning nor their function, but rather the diaphasic level because the diminutive brings a trait of greater informality, as already pointed out by Dressler & Merlini Barbaresi (1994). In fact, in the corpora consulted, diminutives are mainly present in exchanges between people who are in a relationship of familiarity or when the speaker wants to show informality. As Caffi (2007: 100) says, diminutives are in fact accelerators of intimacy that stress "the in-group membership" (Blum-Kulka, 1992: 267):

(22) *Capisco che tu sia molto impegnata, e questo è anche giusto, ma ogni tanto trova un momentino* (ItTenTen20)

'I understand that you are very busy, and that is fair enough, but every now and then find a moment.DIM'

(23) *Ciao, ciao, care signore, sono emerso un attimino, per dar acqua ai miei fiori invasati* […] (ItTenTen20)

'Hello, hello, dear ladies, I have emerged a little (lit. 'an instant.DIM'), to water my potted flowers […]'

As we can see, in (22) and (23) we have respectively a Cxs[APPROX] and Cxs[QUANT] in which the diminutive forms perfectly fit the informal register used by the speakers.

3.2 From quantitative constructions to interactional functions: CxsTN[HEDGE] and CxsTN[ALERT]

In this Section we deal with CxsTN which behave as discourse markers and have procedural meaning, which consists in giving instructions to the recipient about the structure of the text or about how to interpret it.

The first kind of construction functions as a hedge, as we can see from the comparison between an utterance with and without the TN.

(24) a. *allora lo guardiamo scusate lo ascoltiamo un momento/attimo/minuto okay no pe ho fatto bene perché è troppo importante lo ascoltiamo un momento guardando la parte va bene ascoltiamo solo il tema è una ballad* (VoLIP)

'so we'll watch it sorry we listen to it for a moment ok no I did well because it's too important we listen to it for a moment looking at the part all right we just listen to the theme it's a ballad'

b. *allora lo guardiamo scusate lo ascoltiamo Ø okay no pe ho fatto bene perchè è troppo importante lo ascoltiamo Ø guardando la parte va bene ascoltiamo solo il tema è una ballad*

'so we'll watch it sorry we'll listen to it Ø ok no I did well because it's too important we'll listen to it Ø looking at the part all right we'll just listen to the theme it's a ballad'

The example is taken from a lecture during which a professor says to his students that they are going to listen to a ballad and quite clearly, he is not proposing to listen to it for just a moment. In fact, he says in the previous utterance: *ho fatto bene perché è troppo importante* ('I did well because it is too important'). The CxsTN[HEDGE] rather signals to the addressees how the quality of the event is to be interpreted: the original meaning of brevity becomes a metaphor for reduction of effort, levity. In other words, the professor could say "we're going to listen to it without analysing it in depth". This is a case where Dressler & Merlini Barbaresi's (1994) label of [fictive] [non-serious] in its [non-important] component is appropriate. In other words, these constructions delimit the boundaries of the interpretative space of predication and function as hedges (Lakoff 1972). Moreover, the reduction of the scope of predication has consequences on the pragmatic-discursive level, which manifests itself in the attenuation of the force of the illocutionary act (Caffi 2007; Overstreet 2011). The speaker, in fact, reduces his commitment with respect to the truth of the locutionary act from which a manifestation of clear relational vagueness results, also towards his/her addressee.

These constructions can occur with the TNs' diminutive forms, without changing their basic function; and in these cases, the diminutives introduce a meaning of effort's reduction, levity, or proximity between the interlocutors, according to the meaning of predication, as we can see in the following examples:

(25)   *in tarda mattinata e tutto il pomeriggio se mi puoi fare un colpo di telefono così ne parliamo un attimino/momentino/minutino* (VoLIP)

'late morning and all afternoon if you can give me a call so we can talk about it for an instant.DIM/a moment.DIM /a minute.DIM'

(26)   *va avanti un pezzettino dove trova via Larga gira un attimino/momentino/?minutino a destra poi sulla sinistra in via Pantani* (VoLIP)

'go ahead a little bit where you find via Larga turn right an instant.DIM/a moment.DIM /?a minute.DIM then left into via Pantani'

In (25) *ne parliamo un attimino* ('we can talk about it for an instant.DIM') indicates an informal way of talking without a particular commitment by the speakers; in (26) *gira un attimino a destra* ('turn right an istant.DIM') cannot have the same meaning of (25), but indicates a sort of effort's reduction.

Another interactional use of CxsTNs has the function of alerting the interlocutor to what will follow verbally or factually (Bazzanella 1995, Sansò 2020).[7]

(27)   *no no no// un attimo/momento/minuto// voglio vedere chi è* (VoLIP)
'no no no no an instant/moment/minute I want to see who it is'

(28)   *//un attimo/momento/minuto// prego* (VoLIP)
'one moment please'

(29)   *//ah un momento/attimo/minuto// qui la Germania deve essere ancora divisa* (talking about a geographical map) (VoLIP)
'ah one moment/instant/minute here Germany still has to be divided'

(30)   *//aspetta un attimo/momento/minuto// scusa c'è qualcosa che non torna aspetta fermo lì ah fermo ci sei* (VoLIP)

'wait an instant/moment/minute sorry there is something wrong wait there stop there ah stop you are there'

---

[7] The double slash indicates the boundary of a tone unit.

In examples (27)–(30) the CxsTNs clearly have the function of warning, alerting the addressee and soliciting his/her attention. For this reason, these usages are often combined with imperatives: *aspetta* ('wait'), *scusa* ('excuse'), *ascolta* ('listen') which are semantically and discursively equivalent, even though the verb *aspettare* ('wait') covers 50% of all occurrences of this construction in the VoLIP.[8] These constructions, which I label CxsTN[ALERT], are most of the time coextensive with a tone unit, i.e. they are prosodically separate from what precedes and follows.

These constructions are mainly found at the beginning of turns because they have the function of signalling the opening of new discursive developments and thus assume a cataphoric position that introduces textually new elements. This would also seem to confirm that elements expressing subjectivity tend to be in the left periphery of utterances (Traugott 2012: 60). Often these constructions can be used to indicate a change of topic or to introduce a counterargument (Blum-Kulka et al. 1989) with which the speaker wants to signal that the reasoning done up to that point must be changed or has a flaw, e.g. (29)–(30).

CxsTN[ALERT] with diminutives would be entirely acceptable to my native speaker intuition, yet they are not found in the consulted corpora. A tentative explanation for this absence can be derived from the fact that, if the diminutive is an intimacy accelerator or, in any case, signals a desire to get closer to the addressee, it is not suitable for a situation in which a speaker wants to counter-argue or warn his/her addressee about something. In this case the speaker does not feel appropriate to stress the in-group membership.

3.3 Textual uses: CxsTN[FOC]

From a textual point of view, the introduction of a counterargument or the warning that something is going to happen are equivalent to the introduction of a new topic. And indeed, I found cases where CxsTNs seem to have exactly this function. Let's look at this dialogue taken from an oral examination at the university:

---

[8] For the purposes of the present reasoning, it is sufficient to exemplify the construction using always the second person singular of the imperative, but in the corpora, there are also occurrences of the first and second plural and forms with clitic pronouns: *scusami* ('excuse me'.IMP.2PER.SING.CLIT.ACC), scusatemi ('excuse me'.IMP.2PERS.PLU.CLIT.ACC).

(31)   P(rofessor): *bene allora qual è il tema che ti è interessato di più nel secondo modulo*
          P: 'well then what is the topic that interested you most in the second module'

          S(tudent): *allora mh che mi ha che mi è particolarmente piaciuto mh la valutazione e il testing*
          S: 'well mh that I particularly liked mh the evaluation and testing'

          P: *mh mh*

          S: *e mh il discorso riguardante eh mh il laboratorio linguistico il passaggio da eh alt er er e il dalia*
          S: 'and mh the discussion regarding eh mh the language lab the passage from eh alt er er and the dalia'

          P: *ah il mio articulo quello ah d'accordo*
          P: 'ah my article that ah ok'

          S: *sì sì*
          S: 'yes yes'

          P: *va bene senti parliamo un attimo di valutazione*
          P: 'all right listen let's talk *un attimo* about evaluation'

                                            (KIParla)

In the last turn the professor firstly uses the discourse marker *va bene* ('all right') to both acknowledge and close the student's speech and then makes a request that interrupts the student and introduces a new topic *senti parliamo un attimo di valutazione* ('listen let's talk *un attimo* about evaluation') putting it in focus. The focus is a resource available to the speaker which activates the attention of the recipient (Ghesquière 2017), because it represents the point of maximum information or newness or salience and, as such, can only be considered, following Givón (1989), as the extreme of an information continuum whose opposite extreme is the topic. However, there is not a universally accepted definition of focus. Furthermore, Matić and Wedgwood (2013) reveal that the various constructions to which the function of focusers have been attributed show very different meanings and discursive effects (Sturt et al. 2004; König 2007; Ghesquière 2017). However, in Miller's words (2006: 122):

> every concept of focus has to do with giving prominence to constituents and the information they carry, albeit for different reasons – the introduction of new entities or new propositions, the contrast of one entity with another, "exhaustive listing" (one particular entity and no other), or noncontrastive prominence.

Apparently, the final turn in (31) is very similar to examples with CxsTN_[HEDGE] (ex. (25)–(27)), yet if we look at the position of the constructions, we can see some differences. CxsTN_[HEDGE] are usually at the end of the utterance, which in the examples of the corpora coincide with the end of the predication, taking scope over the whole utterance and thus affecting its illocution. In contrast, in (32) the construction occurs between the verb (*parliamo* 'let's talk') and its argument (PP *di valutazione* 'about evaluation'), putting it in foreground. The difference can be seen if we compare the same utterance with the construction in two different positions:

(32) a. *parliamo un attimo* CxsTN_[FOC] *di valutazione*

　　 b *parliamo di valutazione un attimo* CxsTN_[HEDGE]

The different position of *un attimo* produces in (32a) a focus on *di valutazione*, while in (32b) it indicates a more informal way of talking, as in example (24).

Potentially also *momento* ('moment') and *minuto* ('minute') could occur in the same position of *attimo* in (32a), but I did not find any such example in the corpora. In these constructions diminutives are also allowed and frequent, but here their use seems to have a double function:

(33)　　 *vi volevo aggiornare un attimino* CxsTN_[FOC] *su questo disegno di legge che è in discussione* (VoLIP)

　　　　 'I wanted to update you an instant.DIM on this draft law that is being discussed'

In (33) *un attimino* has a twofold function: on a textual level, it focuses the element that follows it (*questo disegno di legge* 'this draft law'), with the effect of drawing attention on it, but on pragmatic level it mitigates the force of the illocutionary act and thus allows the speaker to present what she is proposing as something not overtly demanding. Although paraphrases are always misleading in explaining pragmatic and textual meanings, a try could be 'let's focus on this draft law, but quickly and without too much pedantry'. In other words, while asking those present to focus on the draft law, the speaker is also conveying a low degree of commitment on her part and thus a certain informality.

These two functions expressed by *attimino* produce a clash typical of the rhetorical figure of the antiphrasis which, according to the Oxford English Dictionary, is "a figure of speech by which words are used in a sense opposite to their proper meaning". Antiphrases

are considered to be the juxtaposition of words whose meaning is opposite such as: *What a lucky day!* to mean the exact opposite. In defining antiphrasis, Mortara Garavelli (1988) quotes Mizzau (1984), according to whom antiphrasis is "la forma più aggressiva ed esplicita dell'ironia" ('the more aggressive and explicit form of irony') and is often used just to highlight a comment or a statement, as in the following example: [9]

(34) *assessore oggi a palazzo Valentini c'è aria un attimino tesa forse una crisi alle porte (VoLIP)*

'councillor today there is *un attimino* tense atmosphere at palazzo Valentini (offices of the regional governor of Lazio) perhaps a crisis is just around the corner'

On the one hand, in (34) *un attimino* focalises the constituent *aria tesa* that the speaker considers the sign of a possible political crisis, while, on the other hand, it attenuates the illocutionary force of the statement. This produces an ironic effect, which in actual fact does not mitigate the speaker's remark as a whole.

Other diminutive forms, such as *filino* ('a little', lit. string.DIM) or *tantino* ('a little bit', lit. much.DIM), *pochino* ('a little bit', lit. bit.DIM) can be used as quantitative approximators with an attenuative function and occur in similar contexts with an ironic effect:

(35) B: *senti ma io comunque non posso fare affari con te*
B: 'listen but I can't do business with you anyway'

A: *perchè?*
A: 'why?'

B: *perchè tu sei un tantino* [risata] *un tantino*
B: 'because you are a little [laughter] a little'

A: *un tantino?*
A: 'a little?'

B: *psicolabile*
B: 'psycholabile'   (VoLIP)

In (35) the antiphrastic effect is, if possible, increased by the juxtaposition of the diminutive *un tantino*, which attenuates the illocutionary force of the utterance, and *psicolabile* ('a

---

[9] The antiphrastic figure described here must be distinguished from the ironic meaning linked to some diminutive forms, such as maritino ('hubby'), very common in different languages: Merlini Barbaresi (2015), Gregová (2015).

little.DIM psycholabile'), which is a quite formal technical medical term to define an important pathology. The combination produces again a sort of semantic disorientation, similar to that one of irony.

These uses seem to outline a pattern characterized by the combination of an approximate quantifier in the diminutive form, *un attimino*, *un filino*, *un tantino, un pochino*, and an adjective, whose combination produces an antiphrastic effect. The greater the distance between the attenuation of the diminutive and the meaning of the adjective in that particular semantic frame, the better the effect.

The same antiphrastic effect can be produced by using prosody. In Section 1 we marked the sentence *Ti aspetto da sette minutini* ('I' have been waiting for you for seven minutes.DIM') as unacceptable. In this sentence a definite cardinal numeral *sette* ('seven') is combined with a noun indicating a temporal approximation *minutini* ('minutes.DIM'), producing a semantic clash that would make the sentence unacceptable as an assertion. However, it would become acceptable if the speaker emphasised the word *minutini* by increasing the pitch value and slightly the duration of the syllable carrying the pitch accent *-ti-* (Gili Fivela & Bazzanella 2014: 111). By exaggerating the modulation of fundamental frequency, the speaker can produce a deliberately antiphrastic effect between the numeral *sette* and *minutini*, that in turn produces an ironic utterance.

(36) a.   *//ti aspetto da sette minuTIni//*   statement

      b   *//ti aspetto da sette minu T I ni//*   ironic utterance

This pattern is in line with the fact that in Italian irony can be expressed by prosodic patterns alone or together with other linguistic strategies (Gili Fivela & Bazzanella 2014: 118). In (36) the ironic sense is conveyed by both the prosodic form and the marked use of the diminutive form in a context in which it would not normally be acceptable. In this way, the utterance is anything but a neutral assertion and it rather expresses a reproach or a complaint.

4. Conclusive remarks

Tab. 1 summarizes the analysis so far described, based on the consulted corpora.

Tab. 1: Description of the main features of CxsTN

CxsTN	TNs	Syntax	Functions of diminutive forms	Intentional vagueness
CxsTN[TEMP]	all	NP or PP adjuncts	-	-
CxsTN[APPROX]	all	NP adjuncts	brevity, informality	informational vagueness
CxsTN[QUANT]	?*minuto*	N, A, V modifiers	accelerator of intimacy	informational vagueness
CxsTN[HEDGE]	all	clausal modifier, DM	accelerator of intimacy	relational vagueness
CxsTN[ALERT]	all	clausal modifier, DM	-	-
CxsTN[FOC]	*attimo* only	DM, V/A modifier	antiphrastic effect	-

The analysis of the CxsTNs has confirmed some known facts and highlighted some less known ones. The three TNs can be part of temporal constructions in which they keep all their categorial and syntactic properties. However, they can also be part of constructions in which they lose the categorial property of nominal inflection and the sequence DET+TN (*un attimo*, *un momento*, *un minuto*) can no longer be interrupted by modifiers nor have post-nominal modifiers.

Although I have not carried out a diachronic study and the sequence in which I have presented the CxsNT does not necessarily reflect their temporal development, the synchronic variation I registered in their use presents some of the most common features of the outcomes of grammaticalisation processes (Hopper & Traugott 2002). The TNs can appear in constructions in which they do not have the original temporal meaning but rather a more general meaning of a '(small) indefinite amount', more properly grammatical values of quantifiers or pragmatic values of hedge and alerter.

Constructions with *un attimo* are the only ones that in the consulted corpora occur as focuser (CxsTN[FOC]). Those with *un momento* and *un minuto* may occur in both approximating and hedge constructions, although *un minuto* remains the most anchored to temporal meaning.

Some new CxTNs are expressions of intentional vagueness because they convey informational vagueness, affecting the propositional content of the utterance through temporal (CxsTN[APPROX]) or quantitative approximation (CxsTN[QUANT]) or relational vagueness, affecting the relation between the speaker and/or the illocutionary force of the utterance (CxsTN[HEDGE]).

While the original temporal constructions do not occur in diminutive forms, the departure from temporal meaning makes the use of diminutive forms possible and frequent. As proposed by Grandi (2017), diminutive forms express a deviation by defect from the default values of the base and as far as the TNs considered here are concerned, this results unavoidably in an indeterminate expression. In fact, it is not clear what to consider as the referent of *attimino* ('instant.DIM'), *momentino* ('moment.DIM'), which are not measurable units of time, but also of *minutino* ('minute.DIM'): half minute, a quarter of a minute? This explains why diminutives are admitted in CxsTN[APPROX], [QUANT], [HEDGE], which vehicle intentional vagueness and express by definition subjective meanings and attitudes.

In the consulted corpora, diminutives are not found in CxsTN[ALERT] which express neither informational nor relational vagueness, but on the contrary draw the attention of the interlocutor or introduce a counter argument. Probably the association between a sort of warning and a form that usually attenuates is pragmatically ineffective.

A separate and interesting discussion is needed for CxsTN[FOC] in which, as I said, only *un attimo* and *un attimino* occur. In these constructions there is a clash between the attenuative function of the diminutive and the focalising outcome. The result is an antiphrastic effect that is often used to produce strongly ironic comments.

It is difficult to attribute a constant and unique meaning to diminutives in all the CxsTN because the dimensions of deviation from the base are of different types and depend on the construction in which they occur (Delhay 1999: 83). However, something interesting can be deduced from the use in constructions that allow them as well as from constructions where diminutives are not possible. Diminutives are allowed in all constructions expressing intentional vagueness and their presence conveys a set of semantic values: attenuation, intimacy, informality, desire by the speaker to be closer to the addressee as a courtesy strategy or to make him/her more involved in the conversation. Their meaning, therefore, tends towards the pole of greater subjectivity rather than that of shared intersubjectivity and

precisely because of that it is strongly related to the specific context. In these cases, what Dressler & Merlini Barbaresi (1994) call the fictiveness of diminutives is revealed "which implies reference to the component of the speaker's attitude in the speech event and which naturally inheres in and conforms to the fuzziness of subjective evaluations" (Merlini Barbaresi 2015: 36). If the set of these semantic values does not seem suitable to alert one's addressee, it seems instead very appropriate to create a strong contrast with a meaning that in a given context expresses the opposite of these values. Hence the antiphrasis or irony of certain combinations.

All these considerations suggest further research at both diachronic and synchronic level. A diachronic study is necessary to understand whether the different constructions developed in different periods of time and whether their different semantic and morphosyntactic aspects can be actually attributed to a grammaticalisation process. From a synchronic point of view, it would be useful to assess more in depth, on the one hand, the part played by the subjectivity, for instance eliciting judgements of a sample of speakers, considering also sociolinguistic variables and, on the other, by the context, analysing the semantics of the elements modified by the various constructions.

References

Bazzanella, Carla. 1995. I segnali discorsivi. In Lorenzo Renzi, Gianpaolo Salvi & Anna Cardinaletti (eds.), *Grande grammatica di consultazione*, 225–257. Bologna: il Mulino.

Berthonneau, Anne-Marie. 1989. *Composantes linguistiques de la référence temporelle. Les compléments de temps, du lexique à l'énoncé.* Thèse d'État, Paris VIL.

Blum-Kulka, Shoshana, Juliane House & Gabriele Kasper (eds.). 1989. *Cross-cultural Pragmatics: Requests and apologies.* Norwood (NJ): Ablex.

Blum-Kulka, Shoshana. 1992. The metapragmatics of politeness in Israeli society. In Richard Watts & Ide Sachiko (eds.), *Politeness in Language: Studies in its History, Theory, and Practice*, 255–280. Berlin: Mouton de Gruyter.

Caffi, Claudia. 2007. *Mitigation.* Amsterdam & Oxford: Elsevier.

Channell, Joanna. 1994. *Vague Language.* Oxford: Oxford University Press.

Comrie, Bernard. 1976. *Aspect.* Cambridge: Cambridge University Press.

Cutting, Joan (ed.). 2007. *Vague Language Explored.* Houndmills: Palgrave MacMillan.

Davidse, Kristine, Vandelanotte Lieven & Hubert Cuyckens (eds.), *Subjectification, Intersubjectification and Grammaticalization.* Berlin & New York: De Gruyter Mouton.

Delhay, Corinne. 1999. "Diminutifs" et niveaux de catégorisation. *Faits de langues* 7(14). 79–87.

De Mauro, Tullio. 1982. *Minisemantica.* Roma-Bari: Laterza.

Diewald, Gabriele. 2011. Pragmaticalization (defined) as grammaticalization of discourse functions. *Linguistics* 49. 365–390.

Dressler, Wolfang U. & Lavinia Merlini Barbaresi. 1994. *Morphopragmatics. Diminutives and intensifiers in Italian, German, and Other Languages*. Berlin & New York: Mouton de Gruyter.

Fillmore, Charles J. 2002. Mini-grammars of some time-when expressions in English. In Joan Bybee & Máire Noonan (eds.), *Complex Sentences in Grammar and Discourse*, 31–60. Amsterdam & Philadelphia: Benjamins.

Fleischman, Suzanne & Marina Yaguello. 2004. Discourse markers across languages. Evidence from English and French. In Aida Martinovic-Zic & Carol Lynn Moder (eds.), *Discourse across Languages and Cultures*, 129–147, Amsterdam: Benjamins.

Ghesquière, Lobke. 2017. Intensification and focusing. In Maria Napoli & Miriam Ravetto (eds.), *Exploring Intensification: Synchronic, Diachronic and Cross-linguistic Perspectives*, 189–233. Amsterdam & New York: Benjamins.

Ghezzi, Chiara. 2013. *Vagueness Markers in Contemporary Italian: Intergenerational Variation and Pragmatic Change*. PhD dissertation. Università di Pavia.

Ghezzi, Chiara & Piera Molinelli (eds.). 2014. *Discourse and pragmatic markers from Latin to the Romance languages*. Oxford: University Press.

Giacalone Ramat, Anna. 2015. *Un sacco di* ed altre espressioni di quantità nella prospettiva della grammaticalizzazione. In Maria Grazia Busà & Sara Gesuato (eds.), *Lingue e Contesti. Studi in onore di Alberto M. Mioni*, 583–596. Padova: CLEUP.

Gili Fivela, Barbara & Carla Bazzanella. 2014. The relevance of prosody and context to the interplay between intensity and politeness. An exploratory study on Italian. *Journal of Politeness Research* 10(1). 97–126.

Givón, Talmy. 1989. *Mind, Code and Context*. Laurence Erlbaum: Hillsdale.

Grandi, Nicola. 2017. I diminutivi come marche di attenuazione e indeterminatezza. In Oana-Dana Balaş, Adriana Ciama, Mihai Enăchescu, Anamaria Gebăilă & Roxana Voicu (eds.), *L'expression de l'imprécision dans les langues romanes*, 139–152. Bucharest: Ars docendi - Universitatea din Bucureşti.

Grandi, Nicola & Lívia Körtvélyessy. 2015. Introduction: Why evaluative morphology? In Nicola Grandi & Lívia Körtvélyessy (eds.), *Edinburgh Handbook of Evaluative Morphology*, 3–20. Edinburgh: University Press.

Grandi, Nicola & Lívia Körtvélyessy (eds.). 2015. *Edinburgh Handbook of Evaluative Morphology*. Edinburgh: University Press.

Gregová, Renáta. 2015. Slovakian. In Nicola Grandi & Lívia Körtvélyessy (eds.), *Edinburgh Handbook of Evaluative Morphology*, 296–305. Edinburgh: University Press.

Hopper, Paul J. 1991. On some principles of grammaticalization. In Elisabeth C. Traugott & Bernd Heine (eds.), *Approaches to Grammaticalization*, 17–35. Amsterdam & New York: Benjamins.

Hopper, Paul J. Elisabeth C. Traugott, 2002. *Grammaticalization*. Cambridge: University Press.

Jucker, Andreas H., Sara W. Smith & Tanja Lüdge. 2003. Interactive aspects of vagueness in conversation. *Journal of pragmatics*, 35(12). 1737–1769.

Jurafsky, Dan. 1996. Universal tendencies in the semantics of the diminutive. *Language* 72(3). 533–578.

König, Ekkehard. 2007. *The Meaning of Focus Particles*. London: Routledge.

Körtvélyessy, Lívia & Pavol Štekauer (eds.). 2011. Diminutives and augmentatives in the languages of the world. *Lexis: e-journal in English lexicology* 6. 5–25.

Lakoff, George 1973 [1972]. Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic* 2. 458–508.

Lindblom, Björn. 1990. Explaining phonetic variation: A sketch of the handH theory. In William J. Hardcastle & Alain Marchal (eds.), *Speech Production and Speech Modelling*, 403–439. Dordrecht et al.: Kluwer.

Masini, Francesca. 2016. Binominal constructions in Italian of the N1-di-N2 type: towards a typology of Light Noun Constructions. *Language Sciences* 53. 99–113.

Matić, Dejan & Daniel Wedgwood. 2013. The meanings of focus: The significance of an interpretation-based category in cross-linguistic analysis. *Journal of Linguistics* 49. 127–163. DOI: 10.1017/S0022226712000345.

Merlini Barbaresi, Lavinia 2004. Alterazione. In Maria Grossmann & Franz Rainer (eds.), *La formazione delle parole in italiano*, 264–292. Tübingen: De Gruyter.

Merlini Barbaresi, Lavinia. 2015. Evaluation morphology and pragmatics. In Nicola Grandi & Lívia Körtvélyessy (eds.), *Edinburgh Handbook of Evaluative Morphology*, 32–42. Edinburgh: University Press.

Mihatsch, Wiltrud. 2010. The diachrony of rounders and adaptors: Approximation and unidirectional change. In Gunther Kaltenböck, Wiltrud Mihatsch & Stefan Schneider (eds.), *New Approaches to Hedging*, 93–121. Bingley: Emerald Group.

Mihatsch, Wiltrud. 2016. Type-noun binominals in four Romance languages. *Language sciences* 53. 136–159.

Mihatsch, Wiltrud. 2020. A semantic-map approach to pragmatic markers: The complex approximation/mitigation/quotation/focus marking. In Duarte Oliveira & Romeo Ponce de León (eds.), *Marcadores Discursivos. O Português como Referência Contrastiva*, 137–162, Berlin: Lang.

Miller, Jim. 2006. Focus in the languages of Europe. In Giuliano Bernini & Marcia L. Schwartz (eds.), *Pragmatic Organization of Discourse in the Languages of Europe*, 121–214. Berlin & New York: De Gruyter Mouton.

Miller, Jim & Regina Weinert. 1995. The function of LIKE in dialogue. *Journal of Pragmatics* 23(4). 365–393.

Mortara Garavelli, Bice. 1988. *Manuale di retorica*. Milano: Feltrinelli.

Mizzau, Marina. 1984. *L'ironia. La contraddizione consentita*. Milano: Feltrinelli.

Overstreet, Marianne. 1999. *Whales, Candlelight, and Stuff like That: General Extenders in English discourse*. Oxford: University Press.

Overstreet, Marianne. 2011. Vagueness and hedging. In Gisle Andersen & Karin Aijmer (eds.), *Pragmatics of Society*, 293–318. Berlin & New York: De Gruyter Mouton.

Piantadosi, Steven T., Harry Tily & Edward Gibson. 2012. The communicative function of ambiguity in language. *Cognition* 122(3), 280–291.

Prieto, Victor M. 2015. The semantics of evaluative morphology. In Nicola Grandi & Lívia Körtvélyessy (eds.), *Edinburgh Handbook of Evaluative Morphology*, 21–31. Edinburgh: University Press.

Traugott, Elizabeth C. 1982. From propositional to textual and expressive meanings: Some semantic-pragmatic aspects of grammaticalization. In Winfred P. Lehmann & Yakov Malkiel (eds.), *Perspectives on Historica Linguistics*, 245–271. Amsterdam & Philadelphia: Benjamins.

Traugott, Elizabeth C. 1995. The Role of the Development of Discourse Markers in a Theory of Grammaticalization. Paper presented at the XII International Conference of Historical Linguistics, Manchester.

Traugott, Elizabeth C. 2008. The grammaticalization of NP *of* NP constructions. In Alexander Bergs & Gabriele Diewald (eds.), *Constructions and Language Change*, 21–43. Berlin & New York: Mouton de Gruyter.

Traugott, Elizabeth C. 2012. Intersubjectification and clause periphery. *English Text Construction* 5. 7–28.

Underhill, Robert. 1988. Like is, like, focus. *American Speech* 63. 234–246.

Voghera, Miriam. 2012. Chitarre, violini, banjo e cose del genere. In Anna M. Thornton & Miriam Voghera (eds.), *Per Tullio De Mauro: studi offerti dalle allieve in occasione del suo 80° compleanno*, 341–364. Roma: Aracne.

Voghera, Miriam. 2013. A case study on the relationship between grammatical change and synchronic variation: The emergence of tipo$_{[-N]}$ in Italian. In Anna Giacalone Ramat, Caterina Mauri & Piera Molinelli (eds.), *Synchrony and Diachrony: A Dynamic Interface*, 283–312. Amsterdam & Philadelphia: Benjamins.

Voghera, Miriam. 2017. *Dal parlato alla grammatica*. Rome: Carocci.

Voghera, Miriam. 2019. Numeral constructions in spoken Italian and Spanish: from quantitative approximation to interpersonal relationship. *Normas*. 9(1). 221–233. DOI: 10.7203/Normas.v9i1.16169.

Voghera, Miriam. 2022. Building the reference by similarity: from vagueness to focus. In Hélène Dapote-Vassiliadou & Marie Lammert (eds.), *A Crosslinguistic Perspective on Clear and Approximate Categorisation: A Crosslinguistic Perspective*, 271–298. Newcastle upon Tyne: Cambridge Scholars Publishing.

Voghera, Miriam. 2023. The network of *specie, genere*, *sorta*, *tipo* constructions: From lexical features to discursive functions. In Lieselotte Brems, Kristin Davidse, Inga Hennecke, Alena Kolyaseva, Anna Kisiel & Wiltrud Mihatsch (eds.), *Type Noun Constructions in Slavic, Germanic and Romance Languages*, 351–392. Berlin & New York: De Gruyter Mouton.

Voghera, Miriam & Carla Borges. 2017. Vagueness expressions in Italian, Spanish and English in task-oriented dialogues. *Normas*. 7(1). 57–74. DOI: 10.7203/Normas.7.10424.

Voghera, Miriam & Laura Collu. 2017. Intentional vagueness: A corpus-based analysis of Italian and German. In Maria Napoli & Miriam Ravetto (eds.), *Exploring Intensification: Synchronic, Diachronic and Cross-linguistic Perspectives*, 371–388. Amsterdam & New York: Benjamins.

Miriam Voghera
Dipartimento di Studi umanistici
Università degli Studi di Salerno
Via Giovanni Paolo II
84084 Fisciano (SA), Italy
voghera@unisa.it

Kerstin Majewski

Conference Report:
Symposium "Historical English Word-Formation"
(17–18 February, 2023; Munich, Germany)

The symposium "Historical English Word-Formation" took place on 17 and 18 February 2023 at Ludwig-Maximilians-Universität (LMU) in Munich, Germany. It was organised by the author of the present report and held in memory of the late Prof. Dr. Hans Sauer. The aim was to bring together researchers studying diachronic English word-formation and to showcase current research in this area. Sixteen years ago, it was claimed that diachronic studies of word-formation, especially on the Middle English period, were a desideratum (Kastovsky 2007). The speakers of two plenary sessions and fourteen papers, however, demonstrated that historical word-formation – not least because of the availability of a vast amount of digitised material and large historical corpora – has become a thriving research area. The contributions covered word-formations across a wide variety of text-types and registers, applying different theories and quantitative as well as qualitative methods, and thus offered a great diversity of perspectives. On day one (17 February), five papers focused on the Old English period (c. 700–1100). The second day (18 February) included studies on word-formations in Middle English (c. 1100–1500) as well as Early (c. 1500–1800) and Late Modern English (c. 1800–present).

The event opened with **Mariia Flaksman**'s (LMU Munich) contribution "Onomatopoeic Word-Formations in Old English". Drawing on classifications from previous research, she explained that onomatopoeic words often have an unclear morphological status and discussed whether they should be considered as a means of word-formation at all. Her latest research (Flaksman 2022) demonstrates that most Old English imitative words are found in glosses, but that only a small percentage are imitative by origin (e.g., OE *cracian* > PDE

crack). Furthermore, Old English onomatopoetic words rarely appear in clusters, except in the elegies and riddles of the Exeter Book as well as in heroic poetry (cf. Riddle 24: OE *blǣtan* 'bleat', *beorcan* 'bark', *giellan* 'yell'). A further point she addressed was the disappearance of native (Germanic) onomatopoeia in the history of English: More than half of the Old English imitative words died out (e.g., OE *swinsian* 'to produce a melody'). One of the reasons for their loss might be a change in literary genres and styles. Overall, she argued that the development of onomatopoeic words and their coinages are influenced by a variety of language-internal factors (e.g., changes in the sound-system, in the rules of phonotactics, or in morphology).

The second paper invited the audience to "Dig into Old English Legal Compounds". **Daniela Fruscione** (Goethe University Frankfurt a. M.) and **Letizia Vezzosi** (University of Florence) showed that the Anglo-Saxons, who were the first among the Germanic peoples to codify their laws in the vernacular, created new lexical items in their legal codes through nominal compounding and affixation. The authors presented analyses of various compounds in the law codices of the Kentish kings Æþelbert (*c.* 550–616), Hlothhere († 685) and Eadric († 686/7), Wihtred (*c.* 670–725), and of the West-Saxon kings Alfred the Great (871–899) and Ine († *c.* 726). They detected a profound consistency in the use of compounds and demonstrated that between the earliest texts and later ones, the creation of new compound words gives intriguing insight into legal as well as societal changes in Anglo-Saxon kingdoms. For example, OE *feohfang* can be analysed as a type of *fang* 'taking money', but it came to mean 'the fine for taking money (bribe)'; similarly, *healsfang* was originally a form of corporal punishment ('seizing by the neck') and became 'a legal payment to be paid as a fine'; in a metonymic change, the action to which the law attaches a penalty became the standard word for a fine.

In his plenary speech "The Interface of Old English Dictionaries (IOED) in Database Format: Integrating Derivational Morphology", **Javier Martín Arista** (University of La Rioja) described his ongoing project IOED. This is a relational database which seeks to bring the data in various Old English dictionaries into a comparable form, both at type and at token levels. This is done by relating a headword in one dictionary to its counterparts in other dictionaries and by relating headwords with the same inflectional forms from the same lexical category across dictionaries. An extended version of the database makes it

possible to compare the most widely used corpora of Old English. Javier Martín Arista also discussed the incorporation of morphological relatedness in this database. This was the first of many talks at the Munich symposium that showed how fruitful and necessary corpus linguistics have become for research in historical English word-formation.

The topic of "Old English Occasional Word-Formation: Language Rules, External Influence, and Personal Choices" was explored by **Yekaterina Yakovenko** (Russian Academy of Sciences). On the example of Ælfric's (955–1020) *Grammar* (*Excerptiones de arte grammatica anglice*), which goes back to Latin grammars, she addressed the late Old English word-building processes in linguistic terminology. Ælfric's metalanguage contains loans, loan translations, and periphrastic expressions (e.g., Lat. *modus* – OE *gemet*; Lat. *infinitivus* – OE *ungeendigendlīc*; Lat. *appellativum* – OE *ælc oþer nama*). From a sociolinguistic perspective, these new words arose in a highly educated and cultured environment, which was characterised by cross-linguistic contact (Old English and Latin). Ælfric shows a clear preference for semantic loans and loan translations, which Yekaterina Yakovenko also sees in grammatical treatises in other Germanic languages. Thus, most of his terminology maintains a morphological conformity to the Old English lexical system.

The first day of the symposium concluded with **Mihaela Buzec**'s (Babeş-Bolyai University) contribution "Beyond Rhetoric: The Semantic Components of Kennings and their Role in Aiding Word Retrieval in Oral Old English Poetry". She followed the definition of kenningar in Old English poetry as metaphoric substitutes expressed by nominal compounds. Mihaela Buzec explored the similarities between Old English kenningar functioning as mnemonic devices and characteristics of semantic feature analysis used in the therapy of aphasia and anomia to improve word retrieval in post-stroke patients. Her paper thus brought together the formal and semantic properties of Old English nominal compounding with methods in cognitive linguistics, showing how cognitive linguistics can serve as a framework for the diachronic study of poetic language.

Day two of the symposium was initiated by **Hagen Peukert** (University of Hamburg), who spoke about "Lexical Affix Productivity in the History of English: A Quantitative Approach". He has attempted to collect representational quantitative data on the frequency of more than 300 types of lexical affixes from Old to Present-Day English. He first provided a summary of the computational approaches made so far (e.g., type frequencies to attest

certain affixes with the help of a semi-automatic toolset; a community-based approach for intermediate databases; cf. Peukert 2014; 2016; 2018). Then he discussed the most recent results obtained with the help of the *OED*'s RESTful API (e.g., that the rise in token frequencies of prefixes after the Middle English period suggest that English came to have more characteristics of a prefixing language). One of the central questions was whether Artificial Intelligence can aid identifying individual morphemes in the Middle English lexicon, a period where there was no written Standard but an enormous amount of spelling variations. Overall, his paper showed both the advantages and limitations of corpus and computational linguistics for the analysis of historical language stages.

Another corpus-based study was conducted by **Susanne Lang** (University of Mannheim) on the "Middle English Derivational Suffix *-fien*", summarising the findings of her recently published paper (Lang 2022). In Present-Day English, the suffix *-fy* (< ME *-fien*) is used to form causative verbs (e.g., *falsify* < ME *falsifien* 'to prove (something) to be untrue'). Susanne Lang traced the origin of this function by obtaining data from three Middle English electronic corpora (*The Penn-Helsinki Parsed Corpus of Middle English 2*, *The Parsed Corpus of Middle English Poetry*, and *The Parsed Linguistic Atlas of Early Middle English*). She found that its Middle English derivates pertain to "three main classes […] that represent the three main patterns in which the suffix *-fien* occurs. It either denotes a causative meaning, an abstract causative meaning, or a non-causative meaning" (Lang 2022: 43). So, *-fien* generates not only prototypically causative verbs but also polysemous ones such as *pūrifien* (> PDE *purify*) 'to cleanse; to free the soul from sin or guilt' and non-causative ones like *specifien* (> PDE *specify*).

The importance of sociolinguistic factors, especially of gender, for "[t]he Decline of *-ment*" was investigated by **Tanja Säily** (University of Helsinki) and her colleagues **Martin Hilpert** (University of Neuchâtel) and **Jukka Suomela** (Aalto University). The suffix *-ment* is a borrowing from French. While its productivity reached a peak in the 16th century, it is rather unproductive today. Tanja Säily sought to find out whether gender played a role in the development of different patterns of V-*ment* (cf. Hilpert 2013), drawing on data from the 200-million-word fiction section of the *Corpus of Historical American English* (1810–2009) and using metadata by Öhman et al. (2019). Her results suggest that

women use *-ment* significantly less productively than men, which may relate to women's tendency to employ a more personally involved style of writing (cf. Biber & Burges 2000).

Ursula Lenker (LMU Munich) investigated "Morphologically Marking Epistemicity in the History of English: English *-ly* and the Functional Diversification of English Adverbs". The suffix *-ly* (morphologically, a re-analysis of Old English {*-līc*} (adj.) + {*-e*}) is *the* adverbial ending *per se* in Present-Day English. Yet, the functional reasons for its emergence and establishment have not been explained in much detail. In a case study based on Hans Sauer's editions of *Theodulfi Capitula* (1978) and *The Owl and the Nightingale* (1983), Ursula Lenker compared Old and Middle English strategies for marking epistemicity and evidentiality with those from Early Modern English onwards. In Old and Middle English, stance was mainly expressed by the subjunctive or by impersonal constructions and not by adverbs; the loss of the former was compensated for by a diversification of adverbs. So, Ursula Lenker argues that the development of *-ly* is ultimately connected to signalling speaker attitude.

Paula Rodríguez-Puente (University of Oviedo) gave the second plenary speech at the symposium. It was titled "Register as a Predictor for the Use of Phrasal Verbs: A Diachronic Approach". Phrasal verbs (e.g., *fade away*, *give up*) tend to be associated with spoken language and colloquial registers. This view was challenged by Thim (2012), who argued that the use of these verbs was motivated by the text's subject matter. Aiming at verifying his hypothesis, Paula Rodríguez-Puente investigated a larger set of data (*A Corpus of English Dialogues 1560–1760*, with 1.2-million-words of speech-related text types; and *A Representative Corpus of Historical English Registers 3.1*). She came to the conclusion that register is the main predictor for phrasal verb usage in Early and Late Modern English but that other factors such as changes in style or in genre conventions must also be taken into consideration.

The plenary talk was followed by another corpus-based study of phrasal verbs: **Tara Struik** in collaboration with **Carola Trips** (University of Mannheim) spoke about "Determining the Impact of Verbs Copied from Old French: Towards a Quantitative Analysis of Verbal Prefixes and Verb Particle Combinations in Middle English and Early Modern English". Their research is part of the DFG-project SILPAC ("Structuring the Input in Language Processing, Acquisition and Change"), a lexicon- and corpus-based approach investigating prefix and particle combinations in Middle English and Early Modern English

verbs. This approach seeks to systematically map out the diachronic development of phrasal verbs and detect possible changes in Middle English due to language contact (particularly with French). A preliminary result of their investigation is that French loan verbs and particles did not cause the loss of Germanic prefixes but contributed to the productivity of prefixation. Overall, verbs copied from French seem to have had no decisive impact on the English grammatical system.

Another contribution on Middle and Early Modern English verbs was made by **Michael Bilynsky** (Lviv/Lemberg University): "Hybridity in Middle English: The Case of De-Verbal Families". He studied how inherited bases and suffixes (from Germanic) mix up with borrowed ones (from Latin and French) in Middle English – so-called 'hybrid formations' like PDE *know-able* from OE *cnāwan* + OF *-able* (from Lat. *-ābilis*). Presenting examples from his own database, he raised the question of how to distinguish same-root homogeneous derivatives. The material he investigated were, for example, *-ing* and *-er* de-verbal coinages. According to Michael Bilynsky, hybridity is verifiable by textual evidence and recordable in the lexicography of both word families and historical synonyms.

Next was **Julia Landmann** (University of Basel) with her paper on "Word-Formations Coined from French, Spanish, German, and Yiddish Borrowings in Late Modern English". She presented an approach to lexical borrowing which combined sociolinguistics with cognitive linguistics (cf. Schmid 2018) to explain word-formations – nominal compounds, affixations, conversions, clippings, and blends – based on loans. Her material comprised *c*. 1250 types – the most frequently used loans in Present-Day English retrieved from the *OED*, *OALD*, and *LDOCE*. Julia Landmann investigated the morphological variability of these words in Late Modern English. Also, she looked into the potential loss of semantic connection between the constituents of a given derivative of a borrowing as well as into metonymic or metaphorical processes which led to the creation of a new concept (i.e., the loss of the relation between the constituents). Cognitive aspects relevant to the analysis of the word-formations coined from the various borrowings were also discussed.

The three following papers concentrated on medical and scientific vocabulary. The first was delivered by **Marta Sylwanowicz** (University of Warsaw) and entitled "*Bug doc*, *medico*, *piss-prophet*, *knife happy*, *waterologer*, or *007*: Different Ways of Naming Medical Doc-

tors in English". She examined names and expressions used to refer to medical practitioners in the history of English (from the Middle Ages to today) and drew special attention to non-standard or less formal expressions. She could show that there is evidence of an increasingly critical and negative attitude towards doctors, as many formations include elements of negative sense (e.g., formations with *urine* or *piss*). Marta Sylwanowicz furthermore demonstrated that there is a wide range of word-formation processes involved in the terms for doctors and related professionals. Overall, the data show a trend towards multi-word-units, reduplications, and shortenings.

The two remaining talks focussed on the Royal Society of London's *Philosophical Transactions* which were established in 1665 and which are now "the world's first and longest running scientific journal" (https://royalsocietypublishing.org/rstl/about). First, **Katrin Menzel** (Saarland University) spoke about "Initialisms in Late Modern English Scientific Writing". Her analysis of initialisms in scientific journal articles from 1700–1920 in the *Royal Society Corpus* showed that initialisms for scientific concepts were used only occasionally in this period, referring to persons (honorifics, degrees, members of a society like *F.R.S.* for 'Fellow of the Royal Society') or to institutions (e.g., *R.A.S.* for 'Royal Astronomical Society'). The growing number of shortenings and abbreviations over time can also be seen as a response to the needs of a specialised community of readers. Second, **Magdalena Bator**'s (WSB University in Poznan) paper on "Structural Analysis of Scientific Neologisms – the Case of *Philosophical Transactions*" provided an analysis of the scientific vocabulary first attested in the *Philosophical Transactions* and the different word-formational patterns involved in its creation. Magdalena Bator presented an analysis of the structure (morphology and word-formation), etymology (native or foreign elements), semantics, and survival of the scientific words.

All in all, the Munich symposium stressed the necessity and relevance to engage in historical English word-formation. It showed how diverse this sub-field of English diachronic linguistics has become. Many of the papers also demonstrated how fruitful and crucial it is to integrate findings from corpus linguistics, cognitive linguistics, sociolinguistics as well as from neighbouring disciplines like translation and literary studies.

References

Biber, Douglas & Jená Burges. 2000. Historical Change in the Language Use of Women and Men: Gender Differences in Dramatic Dialogue. *Journal of English Linguistics* 28(1). 21–37. DOI: 10.1177/00754240022004857.

Flaksman, Mariia. 2022. Echoes of the Past: Old English Onomatopoeia. In Sara Lenninger, Olga Fischer, Christina Ljungberg & Elżbieta Tabakowska (eds.), *Iconicity in Cognition and across Semiotic Systems* (Iconicity in Language and Literature 18), 331–350. Amsterdam: Benjamins. DOI: 10.1075/ill.18.

Hilpert, Martin. 2013. *Constructional Change in English: Developments in Allomorphy, Word Formation, and Syntax*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139004206.

Kastovsky, Dieter. 2007. Middle English Word-Formation: A List of Desiderata. In Gabriella Mazzon (ed.), *Studies in Middle English Forms and Meanings* (Studies in English Medieval Language and Literature 19), 41–56. Frankfurt am Main: Lang. https://www.peterlang.com/document/1103010 (accessed 26 February 2023).

Lang, Susanne. 2022. Verifying Causatives: A Corpus-Based Study of the Middle English Derivational Suffix -*fien*. *Mannheim Papers in Multilingualism, Acquisition and Change* 4. 1–56. DOI: 10.25521/mapmac.2022.215.

Öhman, Emily, Tanja Säily & Mikko Laitinen. 2019. Towards the Inevitable Demise of Everybody? A Multifactorial Analysis of -*one*/-*body*/-*man* Variation in Indefinite Pronouns in Historical American English. Paper presented at the 40th Annual Conference of the International Computer Archive of Modern and Medieval English (ICAME 40), Neuchâtel, Switzerland, June 2019. https://tanjasaily.fi/talks/icame40_ohman_et_al_2019.pdf (accessed 26 February 2023).

Peukert, Hagen. 2014. The Morphilo Toolset: Handling the Diversity of English Historical Texts. In Anne Ammermann, Alexander Brock, Jana Pflaeging & Peter Schildhauer (eds.), *Facets of Linguistics: Proceedings of the 14th Norddeutsches Linguistisches Kolloquium 2013 in Halle an der Saale* (Hallesche Sprach- und Textforschung/Language and Text Studies/Recherches linguistiques et textuelles 12), 161–172. Frankfurt a. M.: Lang.

Peukert, Hagen. 2016. Smoothing Derivational Asymmetries in English: In Support of Greenberg's Universal 27. *STUF – Language Typology and Universals* 69(4). 517–545. DOI: 10.1515/stuf-2016-0022.

Peukert, Hagen. 2018. Merging Community Knowledge and Self-Interest to Build Language Resources: Architecture and Quality Management of a Take-and-Share-Approach of Word Annotations. In Manuel Burghardt & Claudia Müller-Birn (eds.), *INF-DH-2018*. Bonn: Gesellschaft für Informatik e.V. DOI: 10.18420/infdh2018-01.

Sauer, Hans (ed.). 1978. Theodulfi capitula *in England: Die altenglische Übersetzung zusammen mit dem lateinischen Text.* Munich: Fink.

Sauer, Hans (ed. and transl.). 1983. *The Owl and the Nightingale/Die Eule und die Nachtigall*. Stuttgart: Reclam.

Schmid, Hans-Jörg. 2018. Ein integratives soziokognitives Modell des dynamischen Lexikons. In Stefan Engelberg, Henning Lobin, Kathrin Steyer & Sascha Wolfer (eds.), *Wortschätze: Dynamik, Muster, Komplexität*. 215–231. Berlin & Boston: De Gruyter. DOI: 10.1515/9783110579963.

Thim, Stefan. 2012. *Phrasal Verbs: The English Verb-Particle Construction and its History* (Topics in English Linguistics [TiEL] 78). Berlin & Boston: Mouton de Gruyter. DOI: 10.1515/9783110257038.

Dr. Kerstin Majewski

Ludwig-Maximilians-Universität München

Department für Anglistik und Amerikanistik

Schellingstr. 3

D–80799 München

Kerstin.Majewski@anglistik.uni-muenchen.de