ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

ARCHIVIO ISTITUZIONALE
DELLA RICERCA

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

The morphological expression of approximation: The emerging simil-construction in Italian

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

Published Version:
Francesca Masini, M. Silvia Micheli (2020). The morphological expression of approximation: The emerging simil-construction in Italian. WORD STRUCTURE, 13(3), 371-402 [10.3366/word.2020.0176].

Availability:
This version is available at: https://hdl.handle.net/11585/802154 since: 2021-02-19

Published:
DOI: http://doi.org/10.3366/word.2020.0176

Terms of use:

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

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

(Article begins on next page)

12 October 2024

This is the final peer-reviewed accepted manuscript of:

Masini, Francesca, and M. Silvia Micheli. "The morphological expression of approximation: the emerging simil-construction in Italian." Word Structure 13.3 (2020): 371-402

The final published version is available online at:

https://doi.org/10.3366/word.2020.0176

Terms of use:

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

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

When citing, please refer to the published version.

The morphological expression of approximation: the emerging *simil-* construction in Italian

Francesca Masini
Alma Mater Studiorum – University of Bologna
francesca.masini@unibo.it

M. Silvia Micheli
University of Milano-Bicocca
maria.micheli@unimib.it

Abstract
This paper contributes to the study of evaluative morphology by investigating an emerging morphological construction in Italian within the framework of Construction Morphology. The schema in question, which contains the string *simil-* (related to the adjective *simile* 'similar') plus a nominal or adjectival base, is analyzed as a newly-created construction that conveys a number of closely-related senses (i.e., fakeness, imitation, resemblance, vagueness, and kin-categorization) revolving around the functional domain of approximation, which has received much less attention than other domains within evaluative morphology. Beside discussing the formal, semantic and usage properties of *simil-* expressions on the basis of corpus data, we propose a constructional network that accounts for their behavior. Finally, we discuss the nature of *simil-* as an affixoid and explore its relationship with other competing (morphological and, more marginally, analytic) strategies in Italian.

Keywords
evaluative morphology, approximation, categorization, constructionalization, prefixes, affixoids, competition, Construction Morphology, Italian.

1. Introduction*

Evaluative morphology is by now a well-established domain of investigation, as testified by the recent volume edited by Grandi & Körtvélyessy (2015). However, not all semantic functions performed by evaluative morphemes have been equally investigated. Indeed, functions such as diminution, augmentation and intensification have been studied quite extensively (cf., among many others, Merlini Barbaresi & Dressler 1994; Jurafsky 1996; Grandi 2002; Schneider 2003; Bakema & Geeraerts 2004; Prieto 2005; Körtvélyessy & Štekauer eds. 2011; Efthymiou 2015; Efthymiou, Fragaki & Markos 2015; Napoli 2017), whereas others, like approximation, have received much less attention.

* This article is the result of close collaboration between the two authors. Exclusively for the purposes of Italian academia, Francesca Masini is responsible for Sections 1, 4.1, 4.2.2, 4.2.3, 4.3, 5, 7, and M. Silvia Micheli for Sections 2, 3, 4.2.1, 6. We are grateful to the audience at IMM18 and four anonymous reviewers for their valuable comments. At various stages, we had the opportunity to discuss about the morphological expression of approximation with Kristel van Goethem, Muriel Norde and Claudio Iacobini: we thank them for sharing their views and insight with us. The usual disclaimers apply.

The fact that research in the latter functional domain 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. Grandi & Körtvélyessy (2015: 9, 11), for instance, use the triad 'approximation/reduction/attenuation' to refer to this functional domain, other relevant terms being 'deintensification' (cf. e.g. Körtvélyessy 2015: 63) or 'non-prototypicality' (cf. e.g. Cúneo 2015: 630). The term 'deintensification' is especially interesting if we consider that Rainer (2015: 1346) includes 'approximation/attenuation' in the wider category 'intensification', which "comprises not only a high degree but all degrees of intensity". Among these terms, 'attenuation' seems to be slightly privileged to refer to 'low degree' in relation to adjectives, especially in the typological literature.[1]
At the same time, approximation has been at the center of interest of other subfields of linguistics, such as pragmatics and discourse studies, where terms like 'mitigation' (cf. e.g. Caffi 2007), 'indeterminacy' (cf. e.g. Bazzanella 2011), 'imprecision' (cf. e.g. Balaş et al. eds. 2017), 'hedging' (Lakoff 1972; Kaltenböck, Mihatsch & Schneider 2010), 'vagueness' (Channell 1994; Mihatsch 2007) or 'intentional vagueness' (cf. Voghera 2012; Voghera & Collu 2017) are more common. These studies, however, have been taking morphological means to convey approximation in little consideration (compared to other strategies such as discourse markers and particles), with the notable exception of diminutive markers used as attenuation strategies or being derived from approximative values (cf. Merlini Barbaresi & Dressler 1994; Merlini Barbaresi 2015; Grandi 2017).
In this paper, we intend to bridge this gap by concentrating on the expression of APPROXIMATION[2] by means of dedicated morphological means (beyond diminutives). We do so by focusing on the Italian language, which displays a number of morphological processes specifically designed to convey APPROXIMATION, especially prefixes/prefixoids (cf. Iacobini 2004a,b):

(1) a. *fra-* (*fraintendere* 'to misunderstand')
b. *intra-* (*intravedere* 'to catch a glimpse of')
c. *pseudo-* (*pseudogiornalista* 'fake journalist', *pseudo-scientifico* 'not really scientific')
d. *para-* (*parafilosofico* 'not really philosophical', *paraletteratura* 'low quality literature')
e. *quasi-* (*quasi-gol* 'almost a goal', *quasi-fidanzata* 'almost a girlfriend')
f. *semi-* (*semi-libertà* 'partial freedom', *semiautomatico* 'semi-automatic')
g. *so-* (*socchiudere* 'to half-close', *sobbalzare* 'to slightly jump')
h. *sub-* (*subsferico* 'roughly spherical', *subumano* 'subhuman')

Prefixes/prefixoids in (1) have different origins: *fra-*, *infra-*, *para-* and *sub-* have a primarily spatial value (e.g. *para-* originally conveys spatial proximity); *pseudo-* conveys already in Greek the meaning 'fake' (e.g., *pseudoprophḗtēs* 'fake prophet'; cf. Italian *pseudomago* 'fake magician'); *quasi* is an adverb meaning 'almost' which can be preposed to nouns and acquires an APPROXIMATIVE

[1] See, for instance, Bauer (2002: 41), who found that 'attenuation' is one of the most common derivation strategies producing adjectives in his 42-language sample. But see Becker (1974), who uses 'approximation' to refer to color terms like French *blanchâtre* 'whitish', as an anonymous reviewer pointed out.
[2] In the remains of this article, we use 'approximation' as a cover term for the whole functional domain associated with attenuation, vagueness, and the other terms introduced in this section. In order to stress that this is just an arbitrary choice, when used in this sense, we use it with small caps (APPROXIMATION, APPROXIMATIVE, etc.). The choice of the term 'approximation' was mainly driven by the observation that, among the many available terms, this seemed to be, overall, the one less associated with a specific (morphological) phenomenon in the international literature and in the Italian tradition, hence the most 'neutral' one.

meaning (e.g. *quasi-sinonimo* 'near-synonym'). Both *para-* and *pseudo-* frequently occur in medical language, where they combine with nouns and adjectives denoting pathologies to create compounds meaning 'showing symptoms similar to those of X' (e.g. *sindrome parainfluenzale* 'flu-like syndrome' or *pseudoartrosi* 'nearthrosis'). In common language, they usually refer to 'something that is fake, lower in quality' (e.g. *para-/pseudo-letteratura* 'low quality literature', *para-/pseudo-scientifico* 'not really scientific'). In some cases, they are used to express an intentionally vague or unclear concept (*paracorteggiatore* 'a sort of admirer', *pseudo-opportunità* 'a sort of opportunity'), often showing a derogatory nuance. Some of these prefixes are obviously more productive than others: for instance, *pseudo-*, *para-* and *semi-* are currently used to create new words, while *fra-*, *intra-*, *so-* have long been considered not productive (Iacobini 2004a: 112, 153; Iacobini 2004b: 88-89). In addition to these prefixal strategies, Italian conveys APPROXIMATION also by means of the suffixes *-oide* '-oid' (e.g. *comunistoide* '(vaguely) in favor of communist ideas') and *-esco* (e.g. *scimmiesco* 'monkey-like, ape-like'), which normally form similarity adjectives, usually bearing a pejorative connotation (cf. Merlini Barbaresi 2004; Wandruszka 2004).
Beside these already recognized (albeit understudied) strategies for expressing APPROXIMATION (cf. also Section 6), Italian displays what we claim to be an emergent morphological construction, so far undescribed, which is formed by attaching the string *simil-* to a base, as exemplified in (2)[3] (examples taken from the CORIS corpus):[4]

(2) a. ***simil**-marsupio*
SIMIL-marsupium
'sort of marsupium/pouch'
c. ***simil**serra* *riscaldata*
SIMILgreenhouse heated
'a sort of heated greenhouse'
b. *freddo* ***simil*** *siberiano*
cold SIMIL Siberian
'Siberian-like cold'

The goal of this article is to offer a descriptive and theoretical account of the properties and emergence of this *simil-* pattern in contemporary Italian by using the tools of Construction Morphology (Booij 2010), and, in doing so, to contribute to the study of the morphological expression of APPROXIMATION. The article is organized as follows. Section 2 outlines the theoretical framework used for the analysis, whereas Section 3 describes some basic facts about *simil-* complex words. Section 4 contains the core of our analysis, based on corpus evidence. Section 5 addresses the status of *simil-* against the background of the literature on affixoids, while Section 6 discusses the issue of the competition between *simil-* and other morphological and analytic strategies in Italian as a promising line for future research. Some concluding remarks follow in Section 7.

[3] *Simil-* is glossed SIMIL throughout the article. Since the orthography varies, as (2) shows (see Section 4.2.1), we always retain the original spelling both in the example and in the glosses.
[4] All examples in this paper are taken either from CORIS or itWaC, unless otherwise specified. See Section 4.1 for details on these two corpora.

2. Theoretical background

In order to analyze the structural and semantic properties of *simil-* we adopt the framework of Construction Morphology (Booij 2010; henceforth CxM), which is briefly introduced in this section. According to CxM, both words and word formation processes are 'constructions' intended as signs, i.e. conventionalized associations of a form and a meaning or function. Since the minimal construction is the word, affixes are not independent entities: the information associated with them is only accessible through the construction in which they occur. Constructions are represented by means of schemas and sub-schemas, differing in complexity and schematicity, connected by means of hierarchical and paradigmatic relations. See for instance Figure 1, which illustrates the portion of constructional network for English prefixation.

a. $< [y[x]_{Xj}]_{Xi} \leftrightarrow [\text{SEM SEM}_j]_i >$

b. $< [\boldsymbol{un}[x]_{\mathbf{A}j}]_{\mathbf{A}i} \leftrightarrow [\textbf{NOT SEM}_j]_i >$

c. $< [un[\boldsymbol{steady}]_A]_A \leftrightarrow [\text{NOT } \textbf{STEADY}] >$ $< [un[\boldsymbol{happy}]_A]_A \leftrightarrow [\text{NOT } \textbf{HAPPY}] >$ $< [un[\boldsymbol{suitable}]_A]_A \leftrightarrow [\text{NOT } \textbf{SUITABLE}] >$ …

Figure 1. Hierarchy for English prefixation (adapted from Masini & Audring 2019: 370).

The higher-level schema in (a) represents the more schematic construction for category-neutral prefixation, which is instantiated by semi-specified constructions like the one in (b), formed by the prefix *un-* and an adjectival variable base. Schema (b), in turn, is instantiated by the fully lexically specified constructions in (c), namely existing *un*-adjectives (e.g. *unsteady*, *unhappy*, *unsuitable*, etc.). The relationship between the higher schema and its instantiations is represented by means of inheritance links: more specified constructions inherit the properties of the overarching construction, spell out their own specifications and may also display overriding features with respect to the mother construction, in line with *default* inheritance.

Like derivation, compounding can be represented in terms of schemas too, as illustrated in Figure 2.

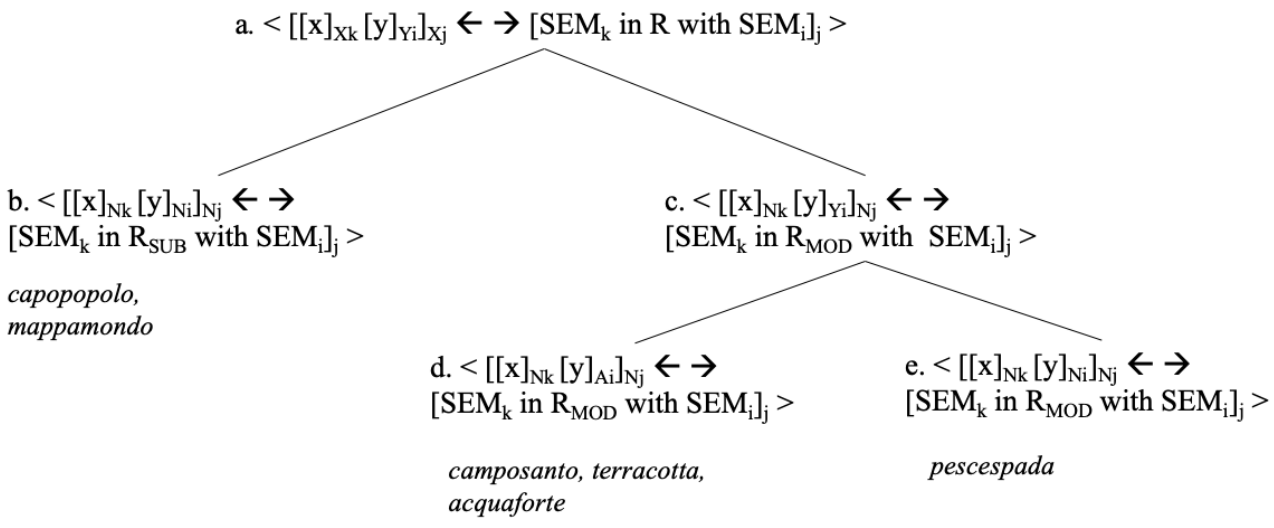

Figure 2. Hierarchy for left-headed compounding in Italian.

The construction in (a) illustrates the general schema for left-headed compounding in Italian, where left-headedness is represented through the category 'X' shared by the leftmost constituent and the whole compound; semantically, the two constituents are linked by a general relation 'R'. This schema is instantiated by two subschemas, (b) and (c), depending on the type of R connecting the two constituents, i.e. subordination (SUB) or modification (MOD).[5] The schema in (b) represents subordinative left-headed compounds, consisting in two juxtaposed nouns (e.g. *capopopolo* 'leader', lit. chief+people; *mappamondo* 'world map', lit. map+world); the schema in (c) is instantiated by two subschemas depending on input categories: Noun-Adjective in (d) (e.g. *camposanto* 'graveyard', lit. field+holy; *terracotta* 'terracotta', lit. earth+cooked; *acquaforte* 'etching', lit. water+strong) and Noun-Noun in (e) (e.g. *pescespada* 'swordfish', lit. fish+sword).[6]

Subschemas are crucial to capture subgeneralizations about morphological processes and to account for the emergence of new constructions, especially for the development of affixes and affixoids from compound constituents (cf. Booij & Hüning 2014). A well-known case is the development of the English suffixes *-hood*, *-dom*, *-ship* (cf. Trips 2009). An intermediate stage is represented by the emergence of affixoids from a free unit occurring within a compound.

Affixoids are compound constituents with an affix-like behavior: an example from Dutch is the adjective *oud* 'old' that occurs in complex words, as their leftmost element, with the meaning 'former', e.g. *oud-leering* 'ex-pupil', *oud-burgemeester* 'former-mayor' (cf. Van Goethem 2008). These complex words represent instantiations of a semi-specified construction composed of *oud* and a base

[5] In this paper, we adopt the classification of compounds into subordinative, coordinative and attributive/appositive, as proposed by Bisetto & Scalise (2005). However, following Arcodia, Grandi & Montermini (2009), we use the term 'modification' to encompass both attribution (e.g. *terracotta* 'terracotta', lit. earth+cooked) and apposition (e.g. *viaggio lampo* 'very short trip', lit. trip+lightning).

[6] Following the traditional classification of Italian compounds (cf. Bisetto & Scalise 2005), we are considering here Noun-Adjective constructions as left-headed compounds. However, we are aware that this classification is problematic. Some scholars (e.g. Gaeta & Ricca 2009: 51-52) consider them as syntagmatic constructions which have undergone a lexicalization process. Also Masini & Scalise (2012: 74) mention these cases in discussing the demarcation between compounds and phrasal lexemes (Masini 2009).

semantically restricted to nouns denoting human beings with a particular social role. When it occurs within this type of complex words, *oud-* shows the more abstract meaning 'former'.
Although the study of the emergence of affixoids has mainly focused on Germanic languages, some cases have also been identified in Romance languages. In Italian, for instance, *mal(e)-* (clearly related to the adverb *male* 'bad') occurs in a large number of complex words expressing pejorative/attenuative or negative values, depending on both the semantic properties of the base and the context, as illustrated in (3) (examples taken from itWaC).

(3) a. *Mio fratello ed io riuscimmo ad ottenere un piccolo scafo, di impossibile e* ***malsicuro*** *equilibrio* […]
'My brother and I managed to get a small hull showing an impossible and unsafe (lit. MAL+safe) balance'
b. *Le persone gravemente* ***malnutrite*** *non riescono a mantenere neanche le funzioni vitali basilari.*
'Extremely undernourished (lit. MAL+nourished) people cannot even maintain basic vital functions.'
c. […] *non si capiva se vivessero volentieri o* ***malvolentieri****, ma comunque vivevano.*
'it was unclear if they lived willingly or unwillingly (lit. MAL+willingly) [in their chaos], but, in any case, they lived'.

In (3a), *mal(e)-* conveys an attenuative value of the base, namely an adjective (*malsicuro* 'unsafe, not completely safe'); in *malnutrito* 'undernourished' (cf. (3b)), *mal(e)-* conveys the information that the act of *nutrire* 'to feed/nourish' is accomplished below an appropriate threshold level. The latter value, namely 'below a threshold', can develop into negation when *mal(e)-* combines with adverbs, as shown in (3c), where *malvolentieri* means 'unwillingly'. As shown by Micheli (2019), the current functions of *mal(e)-* are the result of a gradual process of grammaticalization, whose first signs are recognizable in Old Italian. Particularly, it can be considered as the result of a 'subjectification' process, i.e. "the development of a grammatically identifiable expression of speaker belief or speaker attitude to what is said" (Traugott 1995: 32).
In order to identify the emergence of affixoids from compound constituents, Van Goethem (2010) proposes the following parameters of grammaticalization:[7]

i. 'semantic specialization' (or 'resemantization'), i.e. the rise of a new, more general meaning and/or a metaphorical extension of the meaning displayed by the item as a free lexeme; 'subjectification' (see above) may also be involved;
ii. 'productivity', i.e. the rise of a new subschema that goes along with an increase in terms of productivity;
iii. 'decategorialization' (Hopper & Traugott 1993: 103-113), i.e. the loss of former morphological and syntactic properties (e.g., the loss of inflection in nouns and adjectives);
iv. 'paradigmaticization' (Lehmann 1995: 135), i.e. competition between affixoids and true affixes.

[7] The following parameters are based on several studies on grammaticalization (e.g. Hopper 1991; Hopper & Traugott 1993; Lehmann 1995; Ten Hacken 2000; Marchello-Nizia 2006).

We will return to these parameters in Section 5, where they will be applied to the case of *simil-* in order to verify whether it can be considered an emerging affixoid.

3. Basic facts about *simil-* words

In contemporary Italian, several complex words containing *simil-* as leftmost constituent are attested, as exemplified in (4) (examples taken from itWaC).

(4) a. […] *ci frequentiamo sempre più abitualmente e ci comportiamo*
RECP date.1.PL.PRS always more habitually and REFL behave.1.PL.PRS
come una ***simil-coppia***.
like a SIMIL-couple
'we are dating more and more frequently, and we behave like a pseudo-couple'.
b. […] *ornando di luci* ***simil-natalizie*** *gli alberi.*
adorning of lights SIMIL-Christmas$_{ADJ}$ the trees
'adorning the trees with Christmas-like lights.'

These words are formed by the string *simil-* plus a base which can be either a noun, as in *simil-coppia* 'pseudo-couple', or an adjective, as in *simil-natalizie* 'Christmas-like'. *Simil-*, which is clearly related to the adjective *simile* 'similar, alike',[8] is used here to express APPROXIMATION with respect to the meaning conveyed by the base word: in (4a), *simil-coppia* 'pseudo-couple' refers to an ambiguous relationship between two people that behave like a couple in some respects but are not actually a couple; in (4b), the lights are reminiscent of but not identical to those one finds at Christmas.
As mentioned in the introduction, this strategy is basically not described in reference works. Most notably, *simil-* is not in the list of affixes and formatives in Grossmann & Rainer eds. (2004). The only brief mentions of *simil-* are found in Dardano (1997), who refers to the existence of some *simil-* neologisms in the '90s (e.g. *simil-inglese* 'kind of English', *simil-intervista* 'a sort of interview'), and in Serianni (2004: 591; 2005: 209), who notes that the adjective *simile* 'similar' occurs in compound adjectives, both in rightmost (e.g. *farmaci aspirino-simili* 'aspirin-like drugs', lit. drugs aspirin-similar) and leftmost position (e.g. *quadro simil-reumatico* 'rheumatic-like clinical picture', lit. picture similar-rheumatic). Whereas the former case shows the whole form with inflectional ending (i.e. *simile/simili* 'similar.SG/PL'), the latter displays the phonological reduction (*simil-*) which occurs also in the construction under investigation. In both positions, we obtain compound adjectives referring to diseases/medicines which have symptoms/effects similar to those of a given disease/medicine. In addition to its usage with medical vocabulary, *-simile* rarely occurs with other kinds of bases, but always in health-related contexts, as in illustrated by (5a,b) (from itWac and CORIS, respectively):

(5) a. […] *la sensazione* ***orgasmo-simile*** *dell' eroina.*
the feeling orgasm-similar.SG of_the heroin
'the orgasm-like feeling of heroin'.

[8] Within the morphological construction under investigation, the adjective *simile* always occurs without the final vowel *-e*. As noted by an anonymous reviewer, the final vowel deletion probably allows *simile* to take the form of the minimal prosodic word in Italian (i.e. a foot of two syllables with stress on the first one, cf. Thornton 1996), whose template is crucial in many word formation mechanisms, especially clipping.

b. […] *mescolata a sostanze* ***caffè-simili*** *quali cola e guaranà.*
mixed to substances coffee-similar.PL like cola and guarana
'[…] mixed with coffe-like substances such as cola and guarana'

In both examples, -*simile* conveys a value of APPROXIMATION, close to English -*like*. In (5a), it combines with a base (i.e. *orgasmo* 'orgasm') occurring both in the medical domain and in common language, as in this case (the example is taken from a newspaper article on marijuana). In (5b), it attaches to *caffè* 'coffee' in order to refer to substances that have coffee-like properties; the example is taken from an article on Ecstasy published in a health-related website.
As far as dictionaries are concerned, GRADIT and the Treccani online dictionary[9] mention *simil-* as a formative occurring both in scientific language and in non-scientific language. In the first case, it is claimed to be used to create compound adjectives denoting states/substances similar to other states/substances, e.g. in medical language (*stato simil-tifoso* 'paratyphoid state', lit. SIMIL+typhoid). In the second case, *simil-* is claimed to be used to name imitation/surrogate/low quality products, e.g. *similoro* 'imitation gold' (lit. SIMIL+gold). More specifically, GRADIT and the Treccani dictionary (including the 2008 edition of neologisms) record 14 *simil*-words, illustrated jointly in Table 1.

Table 1. *Simil-* words recorded by GRADIT (G) and the Treccani dictionary (T).

***simil-* complex words**	
similbronzo (G)	SIMIL+bronze
similemofilico (G)	SIMIL+hemophilic
similenzimatico (G)	SIMIL+enzymatic
simil-giovane (T)	SIMIL+young
simil-laburista (T)	SIMIL+Labour
simil-legno (G)	SIMIL+wood
similmarmo (G)	SIMIL+marble
simil-matrimonio (T)	SIMIL+marriage
simil-ormonale (G)	SIMIL+hormonal
similoro (G)	SIMIL+gold
similpelle (G)	SIMIL+leather
similprato (G)	SIMIL+lawn
similproteico (G)	SIMIL+proteinic
simil-seta (G)	SIMIL+silk

Among these examples, we can recognize two leader words[10] which represent the oldest occurrences of *simil*-words, namely: (i) *similoro* 'imitation gold' (metal alloy composed of copper, zinc and tin), attested since 1745; and (ii) *similpelle* 'imitation leather, letherette', attested since 1973 according to GRADIT (but an earlier occurrence in a 1923 text can be found in *Archivio LaStampa*).[11] Notably,

[9] URL: http://www.treccani.it/enciclopedia/elenco-opere/Vocabolario_on_line.
[10] According to Rainer (2013: 152), the term *leader word* was introduced by Spitzer (1925: 588) to indicate a word which serves as a model for the creation of new words and the emergence of a new construction. Although the identification of this phenomenon is far from recent, the notion of leader word and the criteria for identifying it still lack a structured reflection. A first step in this direction is represented by the study by Burdy (2019) on the Latin suffix -*io* in French.
[11] URL: http://www.archiviolastampa.it/.

similoro is a calque from French *similor* 'imitation gold', whose first attestation dates back to 1742, according to Höfler (1981).[12] From a diachronic perspective, the spread of these two *simil-* words can be observed through the Google Books Ngram Viewer graph in Figure 3.

Figure 3. *Similoro* and *similpelle* in the Google Books Ngram Viewer corpus.

The graph in Figure 3 confirms that *similoro* is attested earlier in Italian texts, while *similpelle* emerges in the second half of the 20th century;[13] both words are still used in the present age. Interestingly, both *similoro* and *similpelle* show a 'fake' semantics, since they refer to non-authentic products. Our hypothesis is that a new morphological construction – expressing not just 'fakeness' but also other (closely-related) values of APPROXIMATION – has emerged in present-day Italian. In Section 4 we provide a corpus-based analysis of *simil-*: grounded on data from two corpora of contemporary Italian, both formal and semantic properties will be described and analyzed within the framework of CxM.

4. A corpus-based investigation of *simil-*

4.1 Methodology

Our analysis is based on data extracted from two corpora of contemporary Italian, i.e. the CORIS corpus (Favretti, Tamburini & De Santis 2002), a balanced corpus of written Italian (150M words), and the itWaC corpus (Baroni et al. 2009), a web corpus of 1.5 billion words accessed through the SketchEngine interface (Kilgarriff et al. 2014). Both corpora have been queried for *simil* ngrams and for forms that start with <*simil-*> and <*simil*>, in order to recall *simil-* words in three orthographic variants. i.e. juxtaposition (e.g. *simil pelle*), univerbation (e.g. *similpelle*) and hyphenation (e.g. *simil-pelle*). These procedures ensured high recall but needed a manual check, in order to exclude false

[12] To the best of our knowledge, Höfler (1981) provides the first analysis of French *simili-* as a prefix, whose development apparently displays interesting parallelisms with Italian *simil-*. Although a comparative analysis looks promising and desirable, it goes beyond the scope of the present article.

[13] It should be noted that the graph shows a slight peak before the nineteenth century in the frequency curve of *similpelle*; however, after checking the source, this turned out to be an error due to the OCR software.

positives.
The size of the final datasets with examples extracted from the CORIS and itWaC corpora is illustrated in Table 2.

Table 2. Size of datasets.

CORIS corpus	**itWaC corpus**
219 tokens	1780 tokens
148 types	1456 types

The analysis comprised two stages. First, we carried out an in-depth analysis of data from the CORIS corpus. We analyzed the properties of *simil-* expressions deemed relevant, both formal (i.e. orthography, complexity of the base, category of the base, category of the output) and semantic (i.e. meaning of the base and meaning of the whole expression). Also, we considered a sociolinguistic parameter, namely register of use (i.e. special vs. common language), by checking the usage tag (*marca d'uso*) of the base word available on GRADIT. Afterwards, we performed a more cursory comparison between our findings and the (quantitatively larger) data extracted from the itWaC corpus. The results are detailed in the next subsection (4.2), followed by a discussion and formal analysis (4.3).

4.2 Results

This subsection illustrates the results of our corpus-based analysis, starting from the data we collected from the CORIS corpus, i.e. 219 tokens and 148 types.

4.2.1 Formal properties

As regards orthography, most examples in our dataset are written as one word (38,8% of tokens), like *similpelle* 'imitation leather', or hyphenated (46,1% of tokens), like *simil-esotiche* (lit. SIMIL-exotic.F.PL) 'sort of exotic'.
As for the complexity of the bases, quite expectedly, most bases are simple words (63,9% of tokens, 54,5% of types) or complex words (31% of tokens, 38% of types). However, we also find phrases (5% of tokens, 7,5% of types), see for instance the following examples, where *simil-* has scope over the whole expressions *filmino amatoriale* 'home movie' (6a) and *Humphrey Bogart in "Casablanca"* 'Humphrey Bogart in "Casablanca"' (6b).

(6) a. *simil-filmino amatoriale*
SIMIL-movie.DIM amateurish
'some sort of home movie'
b. *personaggi simil-Humphrey Bogart in "Casablanca"*
characters SIMIL-Humphrey Bogart in "Casablanca"
'characters like Humphrey Bogart in "Casablanca"'

Turning to the category of the base, in most examples, *simil-* attaches to nominals (73,5% of tokens, 64,6% of types) and adjectives (26% of tokens, 34% of types). From a semantic point of view, the nouns used as bases are mostly concrete common nouns (83,9% of tokens, 82,1% of types), especially artifacts (e.g. *camoscio* 'suede' in (7a)) and human beings (e.g. *playboy* 'playboy' in (7b)), but also proper nouns (e.g. *Marilyn* in (7c)) (23% of tokens, 35,8% of types), whereas adjectives are mostly relational (such as *dannunziana* 'of D'Annunzio' in (7d)) (82,1% of tokens, 80% of types).

(7) a. *scarponcini in similcamoscio*
booties in SIMILsuede
'psuedo-suede booties'
b. *un simil-playboy*
a SIMIL-playboy
'a sort of playboy'
c. *una simil-Marilyn*
a SIMIL-Marilyn
'a Marylin-like (character)'
d. *in prosa simil dannunziana*
in prose SIMIL of_D'Annunzio
'in a sort of D'Annunzio-style prose'

Normally, *simil-* attaches to nouns and adjectives to form, respectively, nouns and adjectives (like in (7) above). However, when attaching to nouns (especially nouns denoting humans and proper names), it may result in an adjective-like, attributive construction (7,3% of total tokens), quite similarly to *-like/style* in English (see the translations in the following examples):

(8) a. *un giaccone simil-pompiere*
a jacket SIMIL-fireman
'a fireman-style jacket'
b. *capelli simil cotone idrofilo*
hair SIMIL cotton absorbent
'cotton wool-like hair'
c. *un tempietto di Apollo simil-Delfi*
a temple.DIM of Apollo SIMIL-Delphi
'a little Delphi-style temple of Apollo'

These cases cannot be regarded as nominal appositions, since, if we eliminate *simil-*, the resulting expression is ungrammatical (**un giaccone pompiere*, **capelli cotone idrofilo*, **un tempietto di Apollo Delfi*). Therefore, *simil-* does indeed have category-changing properties under certain circumstances (cf. also Section 4.3).
In addition, in the CORIS corpus we found two examples of *simil-* attaching to different categories, namely a verb (9a) and a preposition (9b). Please note that these uses are definitely marginal and possibly ungrammatical for some speakers (but see Section 4.3).

(9) a. […]*l' aveva capito da come gli* ***similsorrideva***
[…]OBJ have.3.SG.PST understand.PTCP.PST from how to_him SIMILsmiled.3.SG.PST

l' impiegata in camicia color crema.
the employee in blouse color cream
'[he] understood it from the way in which the employee with a creamy blouse was sort of smiling to him'

b. *Bruno Trentin ne sa molto di proteste contro, mezzo-contro,*
Bruno Trentin of_it know.3.SG.PRS much of protests against, half-against,
simil-contro *al governo*
simil-against to_the government
'Bruno Trentin knows a lot about protests against, half-against, SIMIL-against the government'

As for register properties, in quite some cases (23,7% of tokens, 20,4% of types), the base is tagged as "TS" (special language) in GRADIT, and out of these many (48% of tokens, 40% of types) come from medical language.

4.2.2 Semantic properties

Turning now to the semantics of *simil-* expressions, we analyzed and annotated each token from the CORIS corpus according to the specific semantic value they convey. In particular, we could identify five different values, all of which may be regarded as closely related senses of a more general APPROXIMATION function. If Y is the output *simil-* expression, and X is the base to which *simil-* applies, we can define the five values in question as follows:

i. 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; this is the value conveyed by the leader words *similpelle* 'imitation leather' and *similoro* 'imitation gold', among others, and basically coincides with non-authenticity;
ii. IMITATION is the value expressed when: Y refers to an entity that is meant to imitate or reproduce X (without being a genuine X), the difference with (i) being that there is no intent for Y to be taken as an X; this is the value conveyed by expressions such as *simil-Gakona* (lit. SIMIL-Gakona, referring to an exhibition with works of art inspired by the town of Gakona in Alaska, where experiments on electricity are conducted) or *similvita* (lit. SIMIL-life, referring to the virtual life of a videogame character);
iii. RESEMBLANCE is the value expressed when: Y refers to an entity that is merely similar to X without being an X; this is the value conveyed by expressions such as *simil-Amazon* (lit. SIMIL-Amazon, referring to an account that resembles the Amazon account, an Amazon-style account) or *similtropicali* (lit. SIMIL-tropical, referring to beaches that look like tropical ones); items in this class sometimes carry a derogatory nuance, but not necessarily;
iv. VAGUENESS is the value expressed when: Y refers to an entity whose nature is uncertain but somewhat associated to X, possibly as a peripheral member of the category X, having therefore to do with fuzzy category boundaries; this value (corresponding to Voghera's 2012 'intentional vagueness'; cf. Section 1) is conveyed by expressions such as *simil marrone* (lit.

SIMIL brown, referring to a sort of brown) or *similminivan* (lit. SIMIL-minivan, referring a non-prototypical minivan);[14]

v. KIN-CATEGORIZATION is the value expressed when: Y results in a separate (established) entity closely related to X, in that it is defined starting from X and deviating from it; this is the value conveyed by medical expressions such as *simil-influenzale* 'parainfluenzal' and few other common language words such as *simil-matrimonio* (lit. SIMIL-marriage), a neologism referring to civil unions (with derogatory nuance).

Needless to say, these functions are all very close, therefore it was not always easy to assign each item to one class. For instance, the difference between FAKENESS and IMITATION is very slight and we had to resort to the wider context to understand whether the imitation was aimed at being taken for X or not. Another recurrent ambiguity is the one between FAKENESS and RESEMBLANCE: since in few cases we could not solve the puzzle (i.e., we could not guess, even from the wider context, which sense the speaker had in mind), we decided to maintain a FAKENESS/RESEMBLANCE class. For instance, regarding the expression *materiale sintetico similneve* (lit. synthetic material SIMIL-snow), it was unclear whether the synthetic material was intended to be taken as actual snow, or if it was just resembling snow.[15] Table 3 sums up the quantitative results of this semantic annotation.

Table 3. Semantic classification of *simil-* expressions.

Semantic values	**Types**	**Tokens**
VAGUENESS	61	66
RESEMBLANCE	43	49
IMITATION	18	19
KIN-CATEGORIZATION	16	32
FAKENESS	7	49
FAKENESS/RESEMBLANCE	3	4
Total	148	219

As we can see, the more common values, in terms of both types and tokens, are VAGUENESS and RESEMBLANCE, followed by IMITATION (in terms of types only). These first three values display similar figures for types and tokens, meaning that most items are hapaxes and that, consequently, these values are productively conveyed by *simil-*. The FAKENESS value is associated with few types that have many tokens, as expected given that this class contains the two leader words *similpelle* 'imitation leather' and *similoro* 'imitation gold', which are attested since long (cf. Sections 3 and 4.2.3). KIN-CATEGORIZATION shows a similar, although less extreme, situation, since it contains mostly items from medical language that often recur more than once, being rather established terms.

[14] Note that (intentional) VAGUENESS as intended here is different from another value that is also associated with the general domain of APPROXIMATION, namely ATTENUATION. So, for instance, *simil marrone* does not mean 'attenuated/light brown' (a meaning that might be conveyed, instead, by the diminutive form *marroncino*) but 'sort of brown'.

[15] One may argue that a case like this might be ascribed to the IMITATION class too. This is indeed true. However, since IMITATION in our classification stands somehow in-between FAKENESS and RESEMBLANCE, we decided to use these two values to name these ambiguous cases.

Before turning to the next subsection, which offers some quantitative details on the usage of *simil-* expressions, we would like to briefly discuss the nature of *simil-* from a macro-functional point of view. Giving the array of possible values it conveys, *simil-* seems to be at the crossroads of (what are normally regarded as) different domains, namely (new) lexeme creation on the one hand and evaluative morphology on the other. If we construe the creation-evaluation relationship as a continuum rather than a dichotomy, we may say that values such as KIN-CATEGORIZATION and, to a lesser extent, FAKENESS and IMITATION tend towards the creation pole,[16] VAGUENESS definitely belongs to evaluation, whereas RESEMBLANCE stands somewhere in-between, as informally sketched in Figure 4.

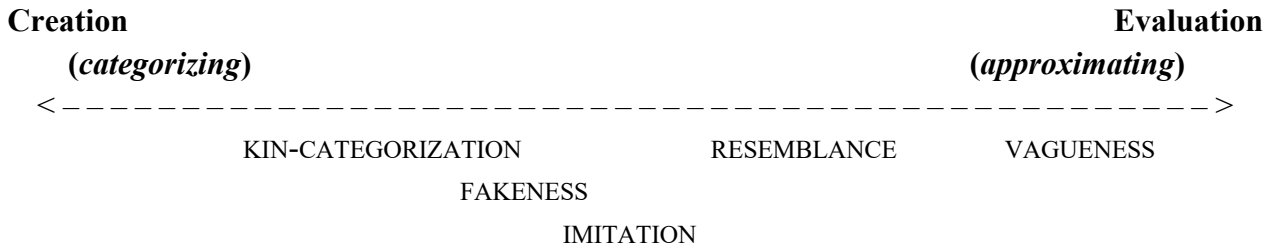

Figure 4. The function of *simil-* between creation and evaluation.

This 'fluid', continuum-like picture also explains why it is sometimes difficult to assign specific occurrences to one class. More in general, the polysemy developed by *simil-* unveils the close relationship between 'approximating' and 'categorizing', two cognitive and linguistic processes that are more closely intertwined than usually thought.

4.2.3 Usage and productivity

The data discussed so far seem to point to a flexible strategy with interesting semantic and formal properties, most notably the possibility to convey a number of closely related functions of APPROXIMATION and to attach to a variety of bases in terms of category (noun and adjectives, but see below) and complexity (words but also phrases).

Speaking of bases, we checked the token frequency of the items to which *simil-* attaches in the CORIS corpus and we found that, apart from very few cases, the vast majority of these items have very low frequency, the hapaxes amounting to 127. Figure 5 shows the bases that occur more than once in the corpus.[17]

[16] FAKENESS seems to be slightly more inclined to lexeme creation than IMITATION probably due to the fact that fake semantics (as described here), and the referents it applies to, are more 'stable' and less ephemeral than mere IMITATION.

[17] Here follow the glosses for the bases in Figure 5, from left to right: 'leather, gold, menstruation, local_policeman, Craxi [Italian politician], influenzal, marriage, hormonal, ecstasy, rubber, marble, hormones, beachrobe, reality_show, foam, schizophrenic, Tarantino, Vina [fictional character], viral'.

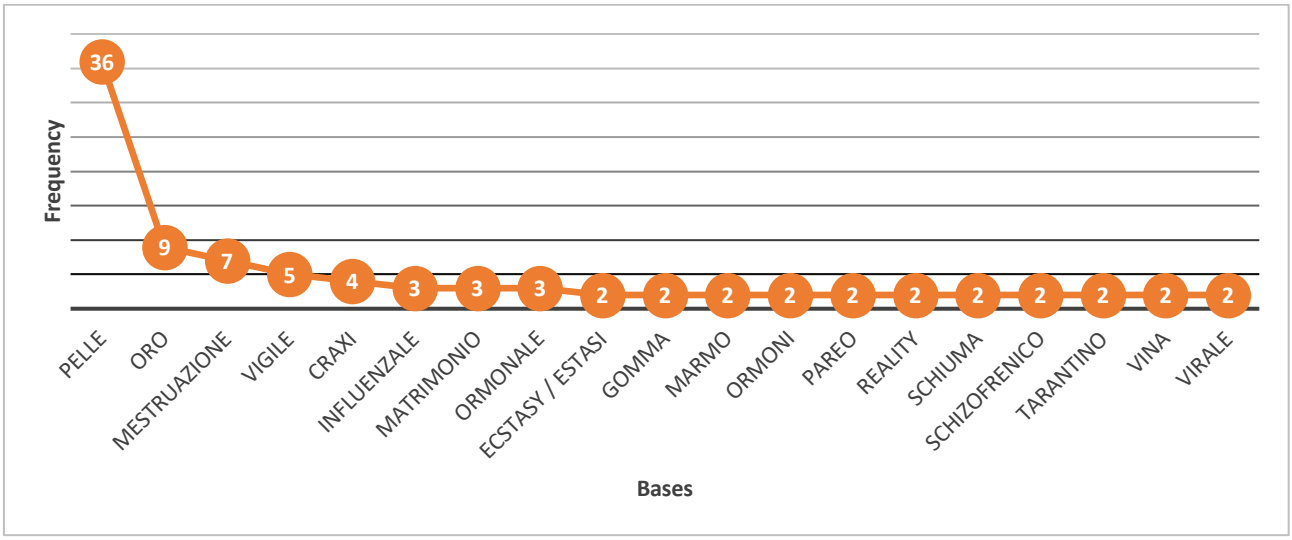

Figure 5. Bases occurring with *simil-* in CORIS (with frequency >1).

As expected, *similpelle* 'imitation leather' and *similoro* 'imitation gold' are the top-ranked expressions in our dataset in terms of token frequency, given that these are the oldest, and hence the most established, occurrences of the *simil-* pattern. Apart from few other cases, among which expressions from the medical special language (*simil-influenzale* 'parainfluenzal', *simil-mestruazione*) and the recent neologism *simil-matrimonio* 'pseudo-marriage' (which is an established item), the items in our dataset are nonce words.

Since the CORIS corpus has been compiled at different times – the first release tracing back to 2001 and being followed by further 'monitor' subcorpora (i.e., texts being constantly added) every 2-3 years – we also checked the micro-diachronic distribution of the *simil-* items within the corpus. The results are showed in Figure 6, which illustrates the distribution of the 219 tokens.

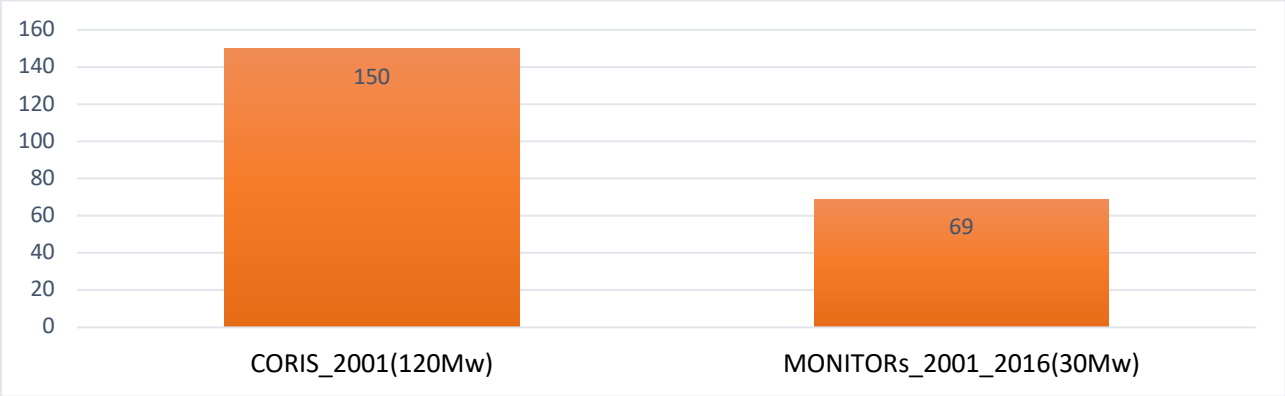

Figure 6. Micro-diachronic distribution of *simil-* words in CORIS.

Considering the different sizes of the two subcorpora (the original 2001 corpus counts around 120M words, whereas the sum of the 'monitor' subcorpora amounts to around 30M words), the presence of *simil-* expressions is more significant in the 'monitor' subcorpora, hence in more recent written texts. The high number of hapaxes and the apparently increasing number of *simil-* expressions in CORIS point to an emergent strategy endowed with a certain degree of productivity. In order to provide more

evidence for this claim we conducted a cursory search for *simil-* expressions in itWaC.
In this (much larger) corpus, automatically compiled from the web, we found 1780 tokens and 1456 types, as already mentioned in Table 2 (Section 4.1). The most frequent items are *similpelle* 'imitation leather' (176 tokens), *simil-influenzale* 'parainfluenzal' (127), *similoro* 'imitation gold' (25), *similcarbonio* (20) 'carbon-like'. On the other hand, the hapaxes amount to 1150.
As for the category of the base, *simil-* attaches mostly to nouns (69,9% of types) and adjectives (30,4% of types), but is also found sporadically with verbs (two occurrences, cf. e.g. (10a)) and adverbs (one occurrence, cf. (10b)).

(10) a. *pure io mi ero messo a* ***simil studiare****...*
also I RFL be.1.SG.PST put.PTCP.PST to SIMIL study.INF
'I had started sort of studying too…'
b. *non hai mai programmato* ***simil-seriamente***
not have.2.SG.PRS never program.PTCP.PST SIMIL-seriously
'You've never programmed semi-seriously'

In general, it seems to us that the findings from itWaC are very much in line with the CORIS data and confirm the general picture provided so far.

4.3 Analysis and discussion

Given the data illustrated in the previous subsections, we hypothesize that *simil-* has been making its way into the Italian lexicon by virtue of the entrenchment of two exemplars – *similoro* 'imitation gold' and *similpelle* 'imitation leather'– that became sufficiently frequent and established. We claim that the entrenchment of these two exemplars led to the emergence of a semi-specified constructional schema, associated with a FAKENESS semantics, where *simil-* is lexically specified and the second slot is a variable that results from a process of abstraction from words such as *oro* 'gold' and *pelle* 'leather'. Once established, this schema, sketched in (11), has motivated the creation of *simil-* words containing nouns belonging to the class of artifacts and natural objects (i.e. the classes to which *pelle* 'leather' and *oro* 'gold' belong), such as *serpente* 'snakeskin' or *marmo* 'marble'.

(11) $< [\textit{simil-}\ [x]_{Nj}\ ]_{Ni} \leftrightarrow [\text{FAKE SEM}_j]_i >$
where x = artifact/natural object

However, this construction appears to have undergone a host-class expansion (Himmelmann 2004) for the variable 'x', which extended from inanimate nouns (artifact/natural object) to human nouns (even proper nouns, following the animacy hierarchy) and adjectives. Sporadically, we found *simil-* attaching to words belonging to other categories, especially verbs, which might hint at a further, ongoing expansion, although evidence is still limited in this respect (see also below).
This class expansion went along with a broadening of meaning, from a purely FAKENESS reading to a wider APPROXIMATION semantics that may be conceived as encompassing the original FAKENESS meaning as well as the other values described in Section 4.2.2, namely: IMITATION, RESEMBLANCE, VAGUENESS, and KIN-CATEGORIZATION.
Since a new form-meaning mapping emerges, we may speak of "constructionalization" as defined by Traugott & Trousdale (2013: 22), namely as "the creation of form$_{new}$-meaning$_{new}$ […] signs", that is

to say of "new type nodes, which have new syntax or morphology and new coded meaning, in the linguistic network of a population of speakers".[18] The new construction can be schematized as follows:

(12) $< [\textit{simil-}\ [x]_{Xj}\ ]_{Xi} \leftrightarrow [\text{APPROXIMATION of SEM}_j]_i >$

This whole process might have been favored both by the lexical semantics of the adjective *simile* 'similar' (which bears a generic meaning of RESEMBLANCE, with no trace of FAKENESS), and by its use in medical special language, where this form is used to create compound adjectives that express new medical concepts by similarity, as already noted (cf. Sections 3 and 4.2.1). In our view, the semantics of 'similarity' (hence, RESEMBLANCE) might have been the key for the transition from the original FAKENESS meaning to APPROXIMATION. Indeed, the link between similative items and the domain of APPROXIMATION has already been recognized (cf. Masini, Micheli & Huang 2018 for a crosslinguistic overview). For instance, Mihatsch (2009) observes that the marker of similative comparison in Romance languages (i.e., French *comme*, Italian *come*, Portuguese *como*, Spanish *como*, equivalent to English 'like') has "developed a derived function as approximation marker" that triggers "loose readings of a lexical expression" (2009: 65-66). According to Mihatsch (2009: 79), "the semantic transition from similative comparison markers to approximation is extremely easy", since "similative comparison is never perfect, but always approximative" (Mihatsch 2009: 70, quoting Haspelmath & Buchholz 1998: 278). In a similar fashion, Bauer et al. (2013: 311-313) note that the suffix *-like* (e.g. *baptismal-like*), when attached to adjectives, seems "to mean not so much 'similar to X' but 'approximating X'", with the approximating reading being derived by inference from the similative one (similar to X > not exactly X > approximating X).
As mentioned above, the more general construction in (12) subsumes five closely related senses revolving around the functional domain of APPROXIMATION (Section 4.2.2). According to our data, the specific meaning *simil-* words acquire partially depends on the type of base ('x') *simil-* attaches to. In particular, we could identify the following tendencies:

i. the FAKENESS meaning is associated with *simil-* words whose bases are concrete nouns denoting artifacts and natural objects, in line with the original construction in (11);
ii. the IMITATION meaning is typically associated with *simil-* words whose bases are concrete nouns denoting humans and locations (occasionally, also artifacts, natural objects, and more abstract nouns, such as *vita* 'life');
iii. the RESEMBLANCE meaning is typically associated with *simil-* words whose bases are concrete nouns denoting humans and artifacts (occasionally, also natural objects and more abstract nouns such as *informazione* 'information');
iv. the VAGUENESS meaning is associated with *simil-* words whose bases are either adjectives or nouns, both concrete and abstract;
v. finally, KIN-CATEGORIZATION is associated with *simil-* words whose bases are either adjectives (belonging to the medical field) or nouns, both concrete and abstract, belonging to a heterogeneous set (among which physiological states and substances).

[18] As the authors observe, constructionalization "is accompanied by changes in degree of schematicity, productivity, and compositionality"; besides, it "always results from a succession of micro-steps and is therefore gradual" (Traugott & Trousdale 2013: 22). Hence, it is closely related to the concept of grammaticalization.

Overall, VAGUENESS is the value which is found with the widest range of possible bases, including bases that are more typically associated with other functions. For instance, we found an occurrence of *simil cemento* (lit. SIMIL cement), used to qualify a grass tennis court, which does not mean 'fake/imitation cement' (as we might have expected given that *cement* is an artifact) but rather 'something vaguely similar to cement'. Whereas the other values have their own niches in terms of semantic class of the base (although there is some overlap), VAGUENESS seems to qualify as some sort of 'default' value which is compatible with basically any base, including adjectives. In fact, adjectival bases occur mostly in the VAGUENESS class, and secondarily in the KIN-CATEGORIZATION class (where, however, we find adjectives related to the medical language only). In addition, the two cases in which the base is not a noun or an adjective (namely, one verb and one preposition) also have a VAGUENESS value.[19] These looser restrictions on the base make the VAGUENESS pattern more available than others, which is testified by the high number of types (and tokens) we found (cf. Table 3, Section 4.2.2). The RESEMBLANCE and IMITATION functions also display some tendency to expand beyond their preferred niches, but to a lesser extent.

How do we account for this variety of meanings in constructionist terms? We propose that, once established, the new construction in (12) has developed its own network of subconstructions, each displaying specific senses and conditions. More specifically, we hypothesize an inheritance hierarchy such as the one illustrated in Figure 7.

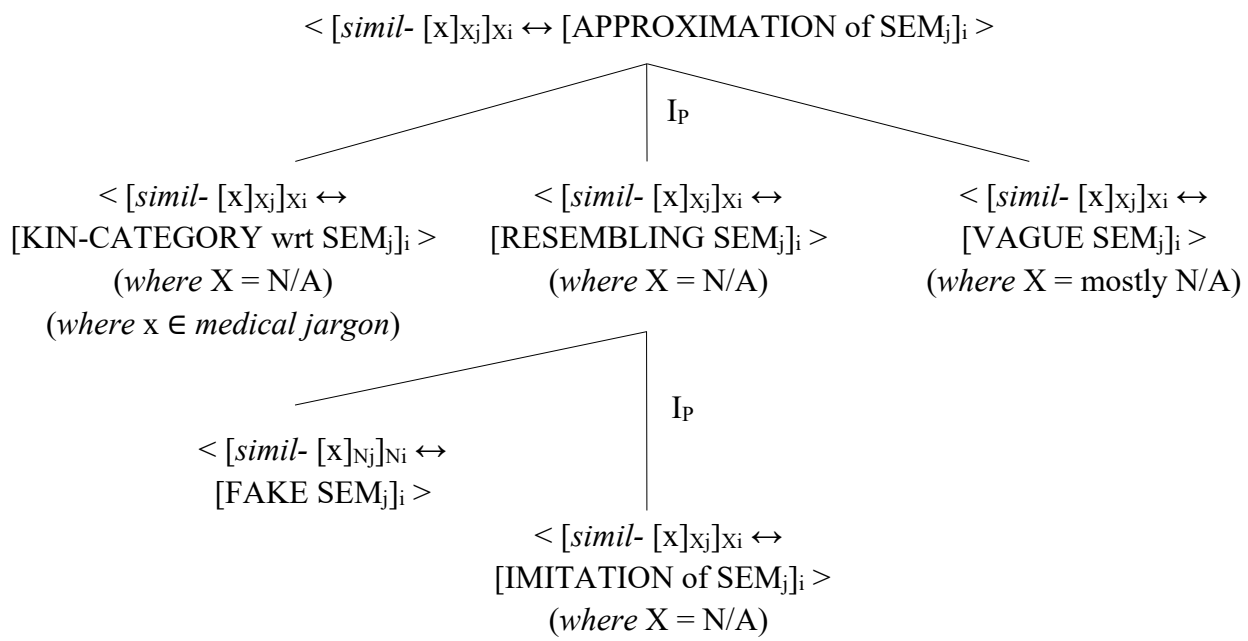

Figure 7. The constructional network for (category-neutral) *simil*-.

The hierarchy is headed by a *simil*- construction with a generic APPROXIMATION meaning (cf. (12)) and underspecified formal properties regarding input and output lexical categories, which is connected – via polysemy inheritance links – to three constructions with more specific formal and semantic properties, namely the subconstructions conveying RESEMBLANCE, VAGUENESS and KIN-CATEGORIZATION.

[19] The same is true of the three instances we found in itWaC in which *simil*- attaches to two verbs and one adverb (cf. Section 4.2.3).

The RESEMBLANCE construction is further connected with two daughter constructions conveying FAKENESS and IMITATION. The latter two meanings can be regarded as more specific senses of RESEMBLANCE: a 'fake' X or something that 'imitates' X is an entity that resembles X in the first place, with additional semantic specifications. These three constructions require nouns as bases (although the RESEMBLANCE and IMITATION constructions also admit adjectives). In most cases, when they attach to nouns they give rise to nouns, but, as already mentioned in Section 4.2.1, in a few cases the whole expression acts like a modifier of a head noun. It is worth noting that these cases all display a RESEMBLANCE semantics, except for two items: one conveys IMITATION (*un tempietto di Apollo simil-Delfi* 'a little Delphi-style temple of Apollo', cf. (8c)) and the other is an ambiguous FAKENESS/RESEMBLANCE case (the already mentioned *materiale sintetico similneve* 'snow-like synthetic material', cf. Section 4.2.2). Hence, the category-changing properties of *simil-* pertains specifically to this branch of the proposed network, in particular to the RESEMBLANCE construction. How do we account for these formations (which are a minority of cases, more precisely 7,3% of total tokens, as reported in Section 4.2.1)? We propose to regard them as instantiations obtained through multiple inheritance: they inherit both from the RESEMBLANCE construction, obviously, and from an abstract construction representing category-changing prefixation, which can be sketched as: < $[y[x]_{Nj}]_{Ai} \leftrightarrow [SEM\ SEM_j]_i$ > (compare the corresponding category-neutral prefixation construction in Figure 1, Section 2).[20] The existence of such a construction is supported by the presence of other category-changing (N>A) prefixes in Italian, such as *anti-* (expressing opposition, e.g. *fari antinebbia* 'fog lights', lit. headlights ANTI+fog) and other spatial or quantitative prefixes (e.g. *multi-*, *bi-*, *post-*), which form invariable adjectives in Italian (cf. Montermini 2008: 197-205). The inheritance from this category-changing prefixation construction guarantees that the denominal *simil-* formations in question are treated as adjectives, and – more importantly – as invariable adjectives, differently from deadjectival *simil-* adjectives (like *simildannunziano* 'of D'Annunzio'), which keep the morphosyntactic properties of the base adjectives (mostly relational in nature).

If we observe Figure 7, we note that the FAKENESS schema (cf. (11)), which we hypothesize to be originally responsible for the emergence of the more general APPROXIMATION construction (cf. (12)), works slightly differently from all the rest, since it only allows for nouns as bases, thus overriding the properties of its mother construction. This 'conservative' trait might speak in favor of our hypothesis. Note that, under this analysis, the FAKENESS construction survives synchronically as a subconstruction of the APPROXIMATION construction itself and, in turn, of the RESEMBLANCE construction. Figure 7 should therefore be interpreted as the synchronic picture of the *simil-* constructional network, not as the diachronic representation of its development, which rather seems to mirror the proposed hierarchy, spreading upwards from FAKENESS to APPROXIMATION through RESEMBLANCE.[21] The analysis proposed in Figure 7 also accounts for apparently unexpected cases such as, for instance, *pattern simil-marmo* (lit. pattern SIMIL-marble) 'marble-like pattern', which carries a RESEMBLANCE reading despite the fact that *marble* is a natural object and hence a good candidate base for the FAKENESS construction. In this case, the 'marble-like' reading is possible because it is licenced directly by the RESEMBLANCE subconstruction, not by the (subordinate) FAKENESS construction.

[20] We are grateful to one of the anonymous reviewers for suggesting to link the *simil-* construction to a more general and abstract construction representing category-changing prefixation in Italian.

[21] In the current paper, we do not pursue a detailed diachronic analysis of *simil-*, which is needed to clarify the precise steps of the development of the various subconstructions proposed in Figure 7.

Finally, the VAGUENESS and KIN-CATEGORIZATION constructions are analyzed as additional subschemas of the APPROXIMATION construction, with their own properties. The latter is strongly associated with bases from the medical jargon (encoded with the preference '*where* x ∈ *medical jargon*'), although we occasionally find items of different sort, like the already mentioned *simil-matrimonio* (lit. SIMIL-matrimonio), which is an established term, like most of the expressions that instantiate this construction. The former, as anticipated above, is slightly more underspecified than other subconstructions, since it appears to be open to bases from different semantic classes (from humans to artifacts to events, etc.) and different lexical categories (but mostly nouns and adjectives, like other *simil-* subconstructions), which enhances the availability of this construction and possibly contributes to the perception of *simil-* as a formative bearing a typically APPROXIMATIVE semantics.

5. What is *simil-*?

To sum up, in previous paragraphs we accounted for the formal and semantic properties shown by *simil-* expressions by hypothesizing a constructional network that captures the more relevant form-meaning associations found in our data. We showed that *simil-* can be considered as a constituent of an emerging morphological construction expressing APPROXIMATION and we regarded the rise of this new construction as a case of constructionalization, as defined in Traugott & Trousdale (2013).
But what is the exact nature of *simil-*? In this section we explore the hypothesis that *simil-* may be regarded as an emerging prefixoid in contemporary Italian by testing the parameters of grammaticalization proposed by Van Goethem (2010) and illustrated in Section 2, namely: (i) semantic specialization or resemantization (plus, possibly, subjectification); (ii) productivity; (iii) decategorialization; and finally (iv) paradigmaticization.
From a semantic point of view, our analysis shows that *simil-* undergoes a broadening of meaning, since it develops a new, more general function of APPROXIMATION which encompasses the original FAKENESS meaning. This resemanticization may have been influenced by the lexical semantics of *simile* 'similar', given the connection between similative items and APPROXIMATION (cf. Section 4.3).
As for the parameter of paradigmaticization, *simil-* turns out to be in competition with true affixes expressing APPROXIMATION, like *pseudo-* or *para-*, as exemplified in (13) (taken from itWaC).

(13) […]*si parla un mix che varia di chilometro in chilometro,*
IMPERS speak.3.SG.PRS a mix that varies of kilometer in kilometer,
mantovano di città, pseudo-bresciano, ***simil-cremonese****, filo-veneto, para-emiliano*
Mantuan of town, pseudo-Brescian, SIMIL-Cremonese, philo-Venetian, para-Emilian
'[…] they speak a mix that varies kilometer by kilometer, [namely] urban Mantuan, pseudo-Brescian, SIMIL-Cremonese, Venetian-oriented, Emilian-like'

Here the conveyed message is that, in the Mantuan area, there are several different dialects resembling the varieties spoken in Brescia, Cremona, the Venetian and Emilian regions. These varieties are named by adding a prefix (*pseudo-*, *simil-*, *filo-* and *para-*) to each relevant toponym. The four prefixes have basically the same function, namely they all indicate that the variety in question resembles the one expressed by the base word (i.e. *bresciano* 'Brescian', *veneto* 'Venetian', etc.) but is not properly ascribable to it.
Note that, compared to APPROXIMATIVE prefixes like *pseudo-*, which has a clearly derogatory

connotation, *simil-* seems to maintain a quite neutral meaning. However, in some cases a pejorative nuance may emerge, as in (14), which hints at some degree of subjectification.

(14) *La scelta è tra una mattonella di* ***simil-pasta*** *e un pollo plasticoso.*
the choice is between a brick of SIMIL-pasta and a chicken plastic-like
'The choice is between a brick of SIMIL-pasta and a plastic-like chicken.'

In (14), *simil-* expresses a pejorative evaluation towards the bowl of pasta, which (shamefully) resembles a brick; the derogatory interpretation is confirmed by the second element, which also shows a pejorative meaning, i.e. *pollo plasticoso* 'a chicken which looks like being made of plastic'.
As far as formal properties are concerned, *simil-* is invariable and cannot be inflected, i.e. the final vowel *-e* never shows. Moreover, we found that *simil-* words are often univerbated orthographically (either with or without a hyphen). These features are signs of decategorialization and loss of autonomy. Finally, we found a high number of hapaxes in both CORIS and itWaC (see Section 4.2.3), which proves that the *simil-* construction is fairly productive.
In accordance with our results, we can conclude that *simil-* can be considered as an emerging and rapidly spreading prefixoid in contemporary Italian.

6. Competition with other morphological and analytic strategies

In the previous section, we observed that *simil-* is in competition with other prefixes to express APPROXIMATIVE values (cf. the 'paradigmaticization' criterion). In what follows, we outline a preliminary overview of the morphological and analytic strategies conveying APPROXIMATION with which *simil-* competes.
Firstly, *simil-* belongs to a set of productive prefixes/prefixoids in present-day Italian which have already been identified by previous works (cf. Iacobini 2004a,b), i.e. *pseudo-*, *para-*, *quasi-* and *semi-*. As the following examples show,[22] all these forms show an APPROXIMATING function.[23]

(15) a. *Ma io ho ancora un po' di aggressività residua*
but I have.1.SG.PRS still a bit of aggressiveness residual
dalla ***pseudo-lite*** *con Giò della sera prima*
from_the PSEUDO-quarrel with Giò of_the evening before
'But I still have some residual aggressiveness from the sort of quarrel with Giò from the night before'.

b. *Il regista aveva spinto fino all' eccesso*
the director have.3.SG.PST push.PTCP.PST until to_the excess
le caratteristiche morbose e ***para-erotiche*** *del fumetto*
the characteristics morbid and PARA-erotic of_the graphic_novel
'The director had pushed the morbid and erotic-like characteristics of the graphic novel to the extreme'

[22] Examples in (15), (16) and (17) were extracted from itWaC.
[23] Each of these forms would require a separate, in-depth semantic and morphological analysis: here we just provide some initial observations.

c. *Il governo Amato approvò un progetto per fare*
the government Amato approve.3.SG.PST a project for make.INF
della vecchia Aurelia una ***quasi-autostrada****: quattro corsie senza pedaggio.*
of_the old Aurelia a QUASI-motorway four lanes without toll
'The Amato government approved a project to turn the old Aurelia street into a quasi-motorway: four lanes, toll-free.'

d. *La giornata più calda di tutta la fiera,*
the day most hot of whole the fair
quasi con un clima ***semi-estivo****.*
almost with a climate SEMI-summery
'The hottest day of the whole fair, almost with a near-summery climate'.

In (15a), *pseudo-lite* refers to a quarrel which is not perceived as a real quarrel (but rather as 'a sort of quarrel') by the speaker. In (15b), the characteristics mentioned by the speaker are not precisely those found in the erotic genre but are reminiscent of it: *caratteristiche para-erotiche* stands for 'erotic-like characteristics' (but not quite so). The *quasi-autostrada* 'almost motorway' mentioned in (15c) is a road that shows many properties of a motorway but not all of them: it has four lanes as a motorway, but it is toll-free (while Italian motorways generally require a toll). Finally, in (15d), the adjective *semi-estivo* 'nearly summery' indicates that the climate is almost as warm as in the summer, but not exactly summery.

Overall, all these prefixes/prefixoids are used to APPROXIMATE. However, some differences emerge which are probably related to the original semantics of each form. *Pseudo-* (originally meaning 'fake') is often used to indicate that something is untrue, false, thus bearing a derogatory interpretation. *Quasi-* 'almost' and *semi-* 'half' (whose original meanings refer to a degree measurement) are used to indicate something that is almost identical to the meaning of the base word, but lacks some (key) properties that would allow it to be considered as a proper member of the category in question. *Para-*, instead, originally expresses a spatial value: when it is used to express APPROXIMATION, it indicates that something is close/related to a given stable category but not properly ascribable to it.

Although these prefixes/prefixoids occur mainly in combination with nouns and adjectives, there are also cases where they attach to verbs, similarly to what observed for *simil-* (cf. Sections 4.2.1 and 4.2.3). In the example in (16), *quasi-cercare* indicates that the action of *cercare* 'to look for' is not carried out completely and adequately but only partially or without much conviction.

(16) […]*36 anni, età che vede parecchi giovani italiani a casa*
[…]36 years, age that see.3.SG.PRS several young Italians at home
dalla mamma a ***quasi-cercare*** *un primo lavoro*
to_the mum to QUASI-look_for.INF a first job
'36 years old, an age at which several young Italians stay at their mother's home sort of looking for a first job'

In addition to the above-mentioned prefixes/prefixoids, which have already been identified in reference works, we suggest to include also *falso-* 'fake', *finto-* 'fake', *mezzo-* 'half' and *non-* 'not', which work similarly to *pseudo-*, *para-*, etc., as shown in the following examples, where they combine with a nominal/adjectival base.

(17) a. *A* *questo* *giro* *evitiamo* *alcuni* *artisti*
at this turn avoid.1.PL.PRS some artists
complimentosi *e* ***falso-cortesi***
ceremonious and FALSO-courteous
'This time, we avoid some ceremonious and fake-courteous artists'.

b. *Per* *me* *Eva* *è* *solo* *una* *delle* *tante* ***finto-pentite*** *al*
for me Eva be.3.SG.PRS just one of_the many FINTO-repentants to_the
solo *scopo* *di* *far* *durare* *il* *suo* *successo.*
only purpose of make.INF last.INF the her success
'For me Eva is just one of many fake-repentants with the sole purpose of making her success last'.

c. *Quando* *applico* *qualsiasi* *delle* *mie* *teorie* *sinistroidi* *e*
when apply.1.SG.PRS any of_the my theories leftist and
mezzo-pacifiste *allo* *Stato* *di* *Israele* […]
MEZZO-pacifist to_the State of Israel
'When I apply any of my leftist and half-pacifist theories to the State of Israel […]'.

d. *Contestualmente* *sta* *per* *essere* *cantierato* *un* ***"non-tram"***
at_the_same_time stay.3.SG.PRS for be.INF build.PTCP.PST a NON-tram
su *gomma* *che* *costa* *il* *doppio* *di* *un* *normale* *filobus.*
on tyre that cost.3.SG.PRS the double of a normal trolleybus
'At the same time, works for a "non-tram" on tyres which costs twice as much as a normal trolleybus are getting to be started'

In (17a), *falso-cortesi* refers to some artists who pretend to be courteous but that the speaker does not consider as such; similarly, in (17b) *finto-pentite* refers to women who (according to the speaker) would pretend to be repentant, without being truly repentant. In (17c), the speaker defines his own theories as *sinistroidi* 'leftist' and *mezzo-pacifiste* 'half-pacifist': both adjectives (conjoined by coordination) show a morphological element expressing APPROXIMATION, i.e. the suffix *-oide* '-oid' (cf. also below) and *mezzo-* 'half', which also convey a derogatory nuance. Finally, in (17d) one can find a bound use of *non-* combining with a noun (*tram* 'tram'): here *non-tram* refers to a sort of trolleybus that shows some properties of trams (it runs on electricity), but at the same time differs from prototypical trams in that it has wheels.

Finto-, *mezzo-* and *non-* also occur as free forms: *finto* is an adjective meaning 'fake', *mezzo* is an adjective meaning 'half', whereas *non* 'not' is an adverb. However, in these examples they are graphically univerbated to their base through a hyphen. Moreover, in *falso-*, *finto-* and *mezzo-*, the final vowel *-o* is homophonous with the linking element that appears in many neoclassical and Adjective-Adjective compounds (e.g. *cardi**o**logo* 'cardiologist', *ital**o**americano* 'Italian-American'), whose function is to contribute to the wordhood of the structures in which they appear. This suggests that the vowel *-o* does not carry the inflectional features 'masculine' and 'singular' (cf. the discussion about decategorialization in Sections 2 and 5).

This brief tour shows that *simil-* is part of a rather large group of APPROXIMATING prefixes/prefixoids with different origin. Whereas *pseudo-*, *semi-* and *para-* are neoclassical prefixoids whose use is well established in special languages (where they convey what we called KIN-CATEGORIZATION, e.g. in medical language: *pseudomembrana* 'pseudomembrane'), *falso-*, *finto-*, *mezzo-* and *non-* are free forms that in the process of acquiring a bound status in common language. As for *quasi-* (an adverb

meaning 'almost' as a free form), its use as a bound form is found in special languages, where it performs a KIN-CATEGORIZING function (e.g. in physics: *quasi-particella* 'quasiparticle'), but it is also spreading in common language, where it seems to convey a RESEMBLANCE/VAGUENESS value (e.g. *quasi-povertà* 'poverty-like condition').

So far, we discussed competing strategies pertaining to the prefixal domain that are category-neutral. However, as discussed in Sections 4.2.1 and 4.3, *simil-* shows category-changing properties: indeed, when *simil-* attaches to nouns, especially within the RESEMBLANCE subconstruction, it may result in adjective-like, attributive constructions (e.g. *giaccone simil-pompiere* 'a fireman-style jacket', lit. jacket SIMIL-fireman), similarly to other Italian prefixes (e.g., *anti-*) and to *-like* or *-style* in English (e.g. *ball-like structure*, *family-style atmosphere*). This property suggests that we should investigate also the competition between *simil-* and items with category-changing capability .

As anticipated in the introduction, in this respect, there are at least three relevant suffixes that are productively used to express APPROXIMATION: (i) the already mentioned *-oide* '-oid' (e.g. *sinistroide* 'leftist'); (ii) *-esco* (e.g. *scimmiesco* 'monkey-like, ape-like', *avvocatesco* 'lawer-like'), which normally creates similative adjectives (cf. Wandruszka 2004); and (iii) *-eggiante* (e.g. *orientaleggiante* 'oriental-like, resembling oriental', *Iphoneggiante* 'Iphone-like, that looks like an Iphone'), which formally corresponds to the present participle of *-eggiare* verbs.[24] The suffix *-oide* (from Greek *-oeidés* 'similar to') is originally (and still) used in special terminology in order to create new terms that are defined by similarity or analogy with what is denoted by the base (e.g. in chemistry: *alcaloide* 'alkaloid'; but see also *prefissoide* 'prefixoid, prefix-like element' in linguistics), a function that we named KIN-CATEGORIZATION. This stresses once more the link between categorization and approximation that we noticed in Section 4.2.2, and that definitely deserves further investigation, also in diachronic perspective.

To conclude this overview, it is worth observing, very preliminarily, that *simil-* and the morphological strategies mentioned above also compete with other, more analytic strategies in Italian conveying APPROXIMATION – and especially (intentional) VAGUENESS – that have lately received much attention. The latter include constructions with taxonomic nouns that turned into approximators, such as *specie/sorta di* 'kind/sort of' or *tipo* 'type' (cf. e.g. Mihatsch 2007; Masini 2016; Voghera 2013, 2017). See for instance the following example, from Voghera (2013: 284):

(18) […] *quadri* *apparentemente* *tutti* *uguali,* *formati* *da* *un* *reticolo* *di* *linee*
paintings apparently all equal formed by a grid of lines
perpendicolari ***tipo*** ***Mondrian***
perpendicular type Mondrian
'apparently identical paintings, formed by a Mondrian-style grid of perpendicular lines'

In this specific context, *tipo* might be easily replaced by *simil-* (*linee perpendicolari simil-Mondrian*) with hardly any change in meaning, but not by other prefixoids like *pseudo-* or *para-* (**linee perpendicolari pseudo/para-Mondrian*), which do not appear to have the same category-changing properties of *simil-* (and *tipo*). Needless to say, a more thorough investigation is needed to understand in which way and to which extent analytic strategies such as these are actually in competition with morphological strategies.

[24] We are grateful to one of the anonymous reviewers for suggesting to add *-eggiante* to the list.

7. Conclusion

In this paper we used the tools of Construction Morphology combined with an empirical methodology to provide a descriptive and theoretical account of an emerging morphological construction expressing APPROXIMATION in contemporary Italian: the *simil-* construction.

By analyzing the morphological and semantic properties of *simil-* words in two corpora of present-day Italian, we found that *simil-* mostly attached to nouns and adjectives but (rarely) also to other lexical categories such as verbs, prepositions and adverbs. Although it normally does not change the category of its base, we noted that, when attaching to nouns, *simil-* may result in an adjective-like, attributive construction, similarly to other prefixes creating invariable adjectives in Italian (e.g. *anti-*, *multi-*, etc.) and to *-like*/*style* in English.

According to our analysis, the APPROXIMATING *simil-* construction developed through two stages: first, a FAKENESS construction with *simil-* emerges through abstraction from two entrenched leader words conveying non-authenticity (*similoro* 'imitation gold' and *similpelle* 'imitation leather'); second, the FAKENESS construction undergoes a process of host-class expansion and broadening of meaning leading to the constructionalization of a more general APPROXIMATING construction. Since the latter is polysemous, in that it displays a number of closely related senses that correlate with specific formal tendencies, we proposed an inheritance hierarchy hosting different subschemas in order to account for this semantic variation. More specifically, we could identify five subschemas carrying the following values: KIN-CATEGORIZATION, FAKENESS (i.e., the 'original' value), IMITATION, RESEMBLANCE, VAGUENESS. This array of meanings conveyed by *simil-* shows that this element seems to be at the crossroads of what are traditionally regarded as two different domains within morphology, i.e. lexeme creation proper and evaluation. Since the most common value in our dataset turned out to be VAGUENESS (followed by RESEMBLANCE), we may conclude that nowadays the *simil-* construction is more strongly associated with the the 'evaluative' side. In addition, *simil-*'s polysemy suggests that categorization and approximation may be regarded as two poles of a continuum (cf. Figure 4), where an intermediate position is occupied by the RESEMBLANCE value, which, by the way, might have been the key value for the transition from the original FAKENESS meaning to APPROXIMATION, in line with the already recognized link between similative items and the domain of APPROXIMATION (cf. Section 4.3).

Finally, we discussed the nature of *simil-*, which proved to meet the criteria proposed by Van Goethem (2010) for the identification of affixoids, namely resemantization, productivity, decategorialization, and paradigmaticization. The latter property led to a (preliminary) overview of the other possible morphological (and, more marginally, analytic) strategies through which one can express APPROXIMATION in Italian. We showed that *simil-* competes with a large set of prefixes/prefixoids – including well recognized ones (such as *pseudo-* or *para-*) as well as possibly emerging ones (such as *finto-* and *mezzo-*) – and with three suffixes (*-oid*, *-esco* and *-eggiante*). Interestingly, similarly to what we found for *simil-*, most of these elements also occur in special languages and have a KIN-CATEGORIZATION function.

This preliminary survey has highlighted relevant issues which are worthy of deeper exploration. First, we still know too little about dedicated morphological markers of APPROXIMATION: more in-depth studies are needed on each of these strategies to unveil their properties and the conditions that rule their competition. Second, the morphological marking of APPROXIMATION constitutes a privileged field of investigation to understand how this complex functional domain is internally organized into

different but related values and what is the relationship between 'approximating' and 'categorizing'. We hypothesize that the type of source from which the marker develops – e.g., 'fake' items (like *pseudo-*), similative and comparative items (like *simil-* itself or *-like* in English), degree items (like *quasi-*), proximity items (like *near-*), taxonomic items (like *kind/sort*) – may play a major role in solving the puzzle. In this respect, the adoption of a wider cross-linguistic perspective may prove to be crucial.

References

Arcodia, Giorgio Francesco & Grandi, Nicola & Montermini, Fabio. 2009. Hierarchical NN compounds in a cross-linguistic perspective. *Rivista di Linguistica* 21(1). 11–33.

Bakema, Peter & Geeraerts, Dirk. 2004. Diminution and augmentation. In Booij, Geert & Lehmann, Christian & Mugdan, Joachim & Skopeteas, Stavros (eds.), *Morphology: An international handbook on inflection and word-formation*, 1045–1052. Berlin: De Gruyter.

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

Baroni, Marco & Bernardini, Silvia & Ferraresi, Adriano & Zanchetta, Eros. 2009. The WaCky wide web. A collection of very large linguistically processed webcrawled corpora. *Language Resources and Evaluation* 43(3). 209–226.

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

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

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

Becker, Hans Ulrich. 1974. *Die approximativen Farbbezeichnungen in den romanischen Sprachen*. Bonn: Rheinische Friedrich-Wilhelms-Universität.

Bisetto, Antonietta & Scalise, Sergio. 2005. The classification of compounds. *Lingue e Linguaggio* 4(2). 319–332.

Booij, Geert. 2010. *Construction Morphology*. Oxford: Oxford University Press.

Booij, Geert & Hüning, Matthias. 2014. Affixoids and constructional idioms. In Boogaart, Ronny & Colleman, Timothy & Rutten, Gijsbert (eds.), *Extending the scope of Construction Grammar*, 77–105. Berlin: De Gruyter.

Burdy, Philipp. 2019. On the importance of leader words in word formation: The popular transmission of the Latin abstract-forming suffix *-io* in French. *Word Structure* 12(1). 42–59.

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

Channell, Joanna Mary. 1994. *Vague language*. Oxford: Oxford University Press.

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

Dardano, Maurizio. 1997. Simil-inglese e finto-spinoso. *Italienisch* 38. 97–100.

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

Gaeta, Livio & Ricca, Davide. 2009. Composita solvantur: Compounds as lexical units or morphological objects? *Rivista Italiana di Linguistica* 21(1). 35–70.

GRADIT = *Grande dizionario italiano dell'uso*, edited by Tullio De Mauro. Turin: UTET, 1999.

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

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

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

Grandi, Nicola & Körtvélyessy, Lívia (eds.). 2015. *The Edinburgh handbook of evaluative morphology*. Edinburgh: Edinburgh University Press.

Grossmann, Maria & Rainer, Franz (eds.). 2004. *La formazione delle parole in italiano*. Tübingen: Niemeyer.

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

Himmelmann, Nikolaus P. 2004. Lexicalization and grammaticization: Opposite or orthogonal? In Bisang, Walter & Himmelmann, Nikolaus P. & Wiemer, Björn (eds.), *What makes grammaticalization: A look from its components and its fringes,* 21–42. Berlin: De Gruyter.

Hopper, Paul J. 1991. On some principles of grammaticization, In Traugott, Elizabeth Closs & Heine, Bernd (eds.), *Approaches to grammaticalization*, 17–35. Amsterdam/Philadelphia: Benjamins.

Hopper, Paul J. & Traugott, Elizabeth Closs. 1993. *Grammaticalization*. Cambridge: Cambridge University Press.

Höfler, Manfred. 1981. Fr. *simili-*: ein Similpräfix? *Romanische Forschungen* 93(1). 23–36.

Iacobini, Claudio. 2004a. Prefissazione. In Grossmann, Maria & Rainer, Franz (eds.), *La formazione delle parole in italiano*, 97–163. Tübingen: Niemeyer.

Iacobini, Claudio. 2004b. Composizione con elementi neoclassici. In Grossmann, Maria & Rainer, Franz (eds.), *La formazione delle parole in italiano*, 69–95. Tübingen: Niemeyer.

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

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

Kilgarriff, Adam & Baisa, Vít & Bušta, Jan & Jakubíček, Miloš & Kovář, Vojtěch & Michelfeit, Jan & Rychlý, Pavel & Suchomel, Vít. 2014. The Sketch Engine: Ten years on. *Lexicography* 1(1). 7–36.

Körtvélyessy, Lívia. 2015. Evaluative morphology and language universals. In Grandi, Nicola & Körtvélyessy, Lívia (eds.), *The Edinburgh handbook of evaluative morphology*, 61–73. Edinburgh: Edinburgh University Press.

Körtvélyessy, Lívia & Štekauer, Pavol (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.

Lehmann, Christian. 1995. *Thoughts on grammaticalization*. Munich: Lincom Europa.

Marchello-Nizia, Christiane. 2006. *Grammaticalisation et changement linguistique*. Brussels: De Boeck-Duculot.

Masini, Francesca. 2009. Phrasal lexemes, compounds and phrases: A constructionist perspective. *Word Structure* 2(2). 254–271.

Masini, Francesca. 2016. Binominal constructions in Italian of the N1-di-N2 type: Towards a typology of light constructions. *Language Science* 53. 99–113.

Masini, Francesca & Audring, Jenny. 2019. Construction Morphology. In Audring, Jenny & Masini, Francesca (eds.), *The Oxford handbook of morphological theory*, 365–389. Oxford: Oxford University Press.

Masini, Francesca & Micheli, M. Silvia & Huang, Shanshan. 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 2018.

Masini, Francesca & Scalise, Sergio. 2012. Italian compounds. *Probus* 24(1). 61–91.

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

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

Merlini Barbaresi, Lavinia & Dressler, Wolfgang U. 1994. *Morphopragmatics: Diminutives and intensifier in Italian, German, and other languages*. Berlin: De Gruyter.

Micheli, M. Silvia. 2019. The diachrony of Italian *mal(e)*-constructions. From compounding to derivation. *Lingue e Linguaggio* 18(1). 91–118.

Mihatsch, Wiltrud. 2007. The construction of vagueness: 'Sort of' expressions in Romance languages. In Radden, Günter & Köpke, Klaus-Michael & Berg, Thomas & Siemund, Peter (eds.), *Aspects of meaning construction*, 225–245. Amsterdam/Philadelphia: Benjamins.

Mihatsch, Wiltrud. 2009. The approximators French *Comme*, Italian *Come*, Portuguese *Como* and Spanish *Como* from a grammaticalization perspective. In Rossari, Corinne & Ricci, Claudia & Spiridon, Adriana (eds.), *Grammaticalization and pragmatics: Facts, approaches, theoretical issues*, 65–91. Leiden: Brill.

Montermini, Fabio. 2008. *Il lato sinistro della morfologia: La prefissazione in italiano e nelle lingue del mondo*. Milano: FrancoAngeli.

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

Prieto, Victor Moises. 2005. *Spanish evaluative morphology: Pragmatic, sociolinguistic, and semantic issues*. Gainesville: University of Florida (Doctoral dissertation).

Rainer, Franz. 2013. Formación de palabras y analogía. Aspectos diacrónicos. In Pujol Payet, Isabel (ed.), *Formación de palabras y diacronía*, 141–172. A Coruña: Universidade da Coruña, Servizo de Publicacións.

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

Rossini Favretti, Rema & Tamburini, Fabio & De Santis Cristiana. 2002. CORIS/CODIS: A corpus of written Italian based on a defined and a dynamic model. In Wilson, Andrew & Rayson, Paul & McEnery, Tony (eds.), *A rainbow of corpora: Corpus linguistics and the languages of the world*, 27–38. Munich: Lincom Europa.

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

Serianni, Luca. 2004. *Composti* (*Medicina*). In Grossmann, Maria & Rainer, Franz (eds.), *La formazione delle parole in italiano*, 590–591. Tübingen: Niemeyer.

Serianni, Luca. 2005. *Un treno di sintomi: I medici e le parole: percorsi linguistici nel passato e nel presente*. Torino: Garzanti.

Spitzer, Leo. 1925. Zur französischen Wortgeschichte. *Zeitschrift für romanische Philologie* 45. 584–592.

ten Hacken, Pius. 2000. Derivation and compounding. In Booij, Geert & Lehmann, Christian & Mugdan, Joachim (eds.), *Morphology/Morphologie: A handbook on inflection and word formation*, vol. 1, 349–360. Berlin: De Gruyter.

Thornton, Anna M. 1996. On some phenomena of prosodic morphology in Italian: Accorciamenti, hypocoristics and prosodic delimitation. *Probus* 8. 81–112.

Traugott, Elizabeth Closs. 1995. Subjectification in grammaticalization. In Stein, Dieter & Wright, Susan (eds.), *Subjectivity and subjectivisation*, 31–54. Cambridge: Cambridge University Press.

Traugott, Elizabeth Closs & Trousdale, Graeme. 2013. *Constructionalization and constructional changes*. Oxford: Oxford University Press.

Trips, Carola. 2009. *Lexical semantics and diachronic morphology: The development of* -hood*,* -dom *and* -ship *in the history of English*. Tübingen: Niemeyer.

Van Goethem, Kristel. 2008. *Oud-leerling* versus *ancient élève*: A contrastive study of adjectives grammaticalizing into prefixes in Dutch and French. *Morphology* 18(1). 27–49.

Van Goethem, Kristel. 2010. The French construction *nouveau* + past participle revisited: Arguments in favour of a prefixoid analysis of *nouveau*. *Folia Linguistica* 44(1). 163–178.

Voghera, Miriam. 2012. Chitarre, violini, banjo e cose del genere. In Thornton, Anna M. & Voghera, Miriam (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 Giacalone Ramat, Anna & Mauri, Caterina & Molinelli, Piera (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 D'Achille, Paolo & Grossmann, Maria (eds.), *Per la storia della formazione delle parole in italiano: Un nuovo corpus in rete (MIDIA) e nuove prospettive di studio*, 277–307. Firenze: Cesati.

Voghera, Maria & Collu, Laura. 2017. Intentional vagueness: A corpus-based analysis of Italian and German. In Napoli, Maria & Ravetto, Miriam (eds.), *Exploring intensification: Synchronic, diachronic and cross-linguistic perspectives*, 373–391. Amsterdam/Philadelphia: Benjamins.

Wandruszka, Ulrich. 2004. Aggettivi di somiglianza. In Grossmann, Maria & Rainer, Franz (eds.), *La formazione delle parole in italiano*, 394–397. Tübingen: Niemeyer.