

M. Silvia Micheli

## The emergence of approximative values in the Italian prefixes *semi-* and *para-*

**Abstract:** This paper examines the diachrony of two neoclassical prefixes in Italian, namely *semi-* (from Latin *sēmi-* ‘half’) and *para-* (from Ancient Greek *pará* ‘close to’), based on data extracted from corpora. The diachronic analysis reveals that until the 20th century, both the prefixes *semi-* and *para-* were mostly used in scientific terminology, where they conveyed their original meanings: degree and spatial proximity respectively. A cursory synchronic analysis based on data extracted from two newspaper corpora shows that in the last decades, *semi-* and *para-* have been productively used to express evaluative values (i.e., attenuation and approximation). The present study enriches the literature on the sources of approximating morphological markers and the relationship between attenuation and approximation, considered here as two distinct values placed along a continuum.

**Keywords:** approximation, attenuation, diachrony, Italian, prefixes, neoclassical elements

### 1. Introduction

This paper contributes to the literature on the morphological expression of approximation by analyzing two prefixes identified by Masini & Micheli (2020) among those morphological elements that express approximation in Contemporary Italian, namely *semi-* and *para-*. Both prefixes are of classical origin and became part of the Italian lexicon through learned loans from Latin and Ancient Greek (e.g., It. *semidio* ‘demigod’ < Lat. *semideus*; It. *parabola* ‘parable’ < Lat. *parabola* < A. Gr. *parabolé*).<sup>1</sup> Particularly, *para-* derives from the Ancient Greek preposition *pará* ‘close to’ (see Bortone 2010: 291; Luraghi 2003: 131–45), while *semi-* originated from Latin *sēmi-* ‘half’ (cf. Ancient Greek *hēmi-* and Sanskrit *sāmi-*; cf. also Italian *emi-* ‘half’), attested in Latin as both compound constituent (e.g., *semianimis*

---

<sup>1</sup> It should be noted that, notwithstanding their classical origin, *semi-* and *para-* are generally considered as prefixes (not as combining forms) within the literature devoted to Italian word-formation (cf. Iacobini 2004: 88, 101), since they express relational meanings.

‘dying’, lit. half + soul) and free form (i.e., the adjective/noun *semis* ‘half’). The two prefixes are used in present-day Italian, both in scientific terminology and everyday language, as respectively exemplified in (1) and (2).

- (1) a. *parapleurite*  
       PARA + pleuritis  
       ‘parapleuritis’
- b. *semiperimetro*  
       SEMI + perimeter  
       ‘half the perimeter of an object or shape’
- (2) a. *diploma para-universitario*  
       certificate PARA + academic  
       ‘university-level certificate awarded by an institution other than a university’
- b. *seminterrato*  
       SEMI + buried  
       ‘basement’

The words in (1) exemplify the use of *para-* and *semi-* in Italian scientific terminologies, specifically in medicine and geometry lexicon. In *parapleurite*, referring to ‘an inflammation affecting the tissues close to the parietal pleura’, *para-* conveys the original value ‘close to, beside’; likewise, *semiperimetro* describes the exact half of a perimeter, consistently with the value of *semi-* attested in Latin (namely, ‘half’). Both prefixes also appear in everyday language, as exemplified by *diploma para-universitario*, in which *para-* denotes ‘comparable to X, but not at the same level’, and *seminterrato* ‘basement’, where *semi-* means ‘partially X’. These values have been already identified in reference studies on Italian prefixation, such as Iacobini (2004: 131, 153); however, as will be explored in more detail throughout the article, *para-* and *semi-* are also used by speakers to express evaluative meanings, i.e., approximation and attenuation. Notably, the emergence of an evaluative meaning in originally locative prefixes (such as *para-*) is also attested in Modern Greek, as demonstrated by Efthymiou, Fragaki & Markos (2015a). However, in contrast with the Italian prefix, Modern Greek *para-* conveys an intensifying value (‘excess’), such as in *para-cimáme*, meaning ‘to oversleep’ (cf. also *iper-* in Efthymiou, Fragaki & Markos 2015b).

The goal of this study is twofold. On the one hand, it investigates the emergence of evaluative meanings in two neoclassical prefixes through a diachronic analysis based on corpus

data. In doing so, it enriches the literature on the sources of approximating morphological markers. On the other hand, this study deepens the semantic analysis of the domain of approximation by analyzing the relationship between approximation and attenuation: notably, the two values are frequently considered a single semantic function within the literature on evaluative morphology, as testified by the terminology used to refer to them, e.g., “attenuation/approximation” in Grandi (2002, 2017) and Rainer (2015),<sup>2</sup> and “approximation/reduction/attenuation” in Grandi & Körtvélyessy (2015: 9, 11). The analysis of the semantic functions performed by the prefix *semi*- will contribute to the discussion on the boundaries between approximation and attenuation.

The article is organized as follows. Section 2 discusses the data extraction methodology and the parameters of analysis. The results of the analysis are presented in Section 3; the diachronic analysis of each prefix is supplemented with a cursory investigation of its use in Contemporary Italian. In Section 4 the results of the analysis are discussed and some perspectives for future research are suggested.

## 2. Methodology

This investigation is mostly based on data extracted from CODIT (i.e., Corpus Diacronico dell’Italiano ‘Diachronic corpus of Italian’; see Micheli 2022a), i.e., a thirty million word diachronic corpus of written Italian that covers a period from the earliest attestations of the language to 1947.<sup>3</sup> The corpus is structured into five subcorpora (see Tab. 1), corresponding to five historical periods comprised between crucial dates for the history of Italian language and literature (cf. the periodization proposed by D’Achille 1990 and adopted in MIDIA, Iacobini, De Rosa & Schirato 2017). Each subcorpus includes texts belonging to six genres: essays, literary prose, poetry, letters, scientific texts, and theatre; the only exception is the first subcorpus, which does not include scientific texts.

---

<sup>2</sup> Particularly, according to Rainer (2015: 1346), the “approximative/attenuative” function does not represent an autonomous value, but can be ascribed to “intensification”, which comprises not only a high degree but all degrees of intensity.

<sup>3</sup> The corpus is hosted by the Czech National Corpus website; it can be searched through the KonText interface available at this link: <https://www.korpus.cz/kontext/query?corpname=codit> (accessed 3 February 2023).

**Tab. 1:** CODIT: Structure and size

	1 (13 <sup>th</sup> c.–1375)	2 (1376–1532)	3 (1533–1691)	4 (1692–1840)	5 (1841–1947)
Essays	1,545,178	575,328	1,914,096	1,057,300	1,167,999
Letters	28,023	1,120,431	982,128	1,582,788	1,312,679
Poetry	945,047	1,674,318	1,257,071	856,507	1,350,275
Literary prose	1,393,433	1,585,877	1,472,700	1,501,061	1,808,785
Scientific texts	0	536,028	641,160	719,599	647,945
Theatre	65,537	410,363	465,163	449,978	428,717
	3,977,218	5,902,345	6,732,318	6,167,233	6,716,400

Since the corpus contains raw texts, it has been queried through the Kontext interface for forms that start with <para> and <semi>. A manual check was conducted to exclude false positives (e.g., *paradiso* ‘paradise’, and *semina* ‘seeding’); finally, the selected forms were manually lemmatized. Tab. 2 illustrates the number of types and tokens of the *semi*- and *para*- words collected for each subcorpus.

**Tab. 2:** Dataset of words containing *semi*- and *para*- extracted from CODIT

	1	2	3	4	5
<i>Semi</i> -	16 tokens; 3 types	39 tokens; 3 types	281 tokens; 6 types	56 tokens; 23 types	84 tokens; 30 types
<i>Para</i> -	29 tokens; 3 types	75 tokens; 4 types	264 tokens; 6 types	446 tokens; 7 types	323 tokens; 9 types

The last part of the investigation is devoted to a cursory analysis of the prefixes *para*- and *semi*- in present-day Italian, based on data extracted from the *la Repubblica* corpus (see Baroni et al. 2004), comprising newspaper articles published from 1985 to 2000, and from a more recent journalistic corpus, namely the Timestamped JSI web corpus (2014–2016), searched through the Sketch Engine interface. The analysis was based on a sample of 100 occurrences of the two prefixes extracted from each corpus; given the large size of the two corpora, the analysis was limited to cases where the prefix is linked to the base through the hyphen to automatically extract words wherein the strings *para*- and *semi*- represent a true prefix and not a false positive. Table 3 illustrates the number of types and tokens of the *semi*- and *para*- words extracted from the two corpora.

**Tab. 3:** Dataset of words containing *semi*- and *para*- extracted from the two corpora of Contemporary Italian

	<i>la Repubblica</i> corpus (1985–2000)	Timestamped JSI web corpus (2014–2016)
<i>Semi</i> -	100 tokens; 72 types	100 tokens; 70 types
<i>Para</i> -	100 tokens; 63 types	100 tokens; 69 types

The dataset obtained has been analyzed according to the following parameters. From a formal point of view, the complexity of the base and the lexical category of the base and of the output have been considered. Semantically, the values conveyed by *para*- and *semi*- have been investigated through a qualitative analysis of the context of each. In the diachronic analysis based on CODIT, the textual genre in which the words are attested has been also taken into account.

### 3. Analysis

#### 3.1 Analysis of the prefix *semi*-

Although not having particularly high-frequency values, *semi*- is attested in all subcorpora of CODIT.<sup>4</sup> The following table illustrates the distribution of types, tokens, and hapaxes among the five subcorpora.

**Tab. 4:** Type/token frequencies and the number of hapaxes of words including *semi*- within CODIT

<i>semi</i> -	1 (13th c.–1375)	2 (1376–1532)	3 (1533–1691)	4 (1692–1840)	5 (1841–1947)
Types	3	3	6	23	30
Tokens	16	39	281	56	84
Hapaxes	1	1	0	16	21

As shown in Table 4, the type frequency of words containing *semi*- increases over time. More specifically, in the first two subcorpora, *semi*- only occurs in learned loanwords from Latin, where it means ‘half’ (i.e., *semicircolo* ‘half circle’ from Lat. *semicirculus*) or ‘partially/partial’ (i.e., *semidio* ‘demigod’ and *semivivo* ‘unconscious’ from Lat. *semivivus*). The latter value is exemplified in (3).

<sup>4</sup> It should be noted that CODIT also records two words containing *emi*-, namely *emicrania* ‘migraine’ and *emisfero* ‘hemisphere’; they are learned loanwords from Ancient Greek (through Late Latin) whose structure reflects that shown by compounds in Ancient Greek made up of two bound forms.

(3)	<i>E</i>	<i>co'</i>	<i>pugni</i>	<i>stretti,</i>	<i>dette</i>	<i>queste</i>	<i>parole,</i>
	And	with	fists	clenched	say.PTCP.PST	these	words
	<i>cadde</i>	<b><i>semiviva</i></b>	<i>nelle</i>	<i>loro</i>	<i>braccia [...]</i>		
	fall.3SG.PST	SEMI + alive	in the	their	arms		

‘With clenched fists, after saying these words, she fell unconscious in their arms’. (Boccaccio, *Filocolo*, 1336–38)

At this stage, the prefix *semi-* cannot yet be considered a productive prefix since the only attested words containing it are all loanwords from classical languages. This scenario partly changes in the third subcorpus (1533–1691), where *semi-* occurs in three (very frequent) terms belonging to the field of geometry, which are attested only within the subcorpus of scientific texts; these terms include *semidiametro* ‘semi-diameter’ (200 occ.), *semicerchio* ‘half-circle’ (29 occ.), *semiparabola* ‘id.’ (21 occ.). The high value of token frequency for this period is due to the high token frequency of *semidiametro* within two specific texts by Galileo Galilei, namely *Due lezioni intorno alla figura, sito e grandezza dell’Inferno di Dante* (1588) and *Dialogo sopra i due massimi sistemi del mondo* (1632). The use of the prefix *semi-* in combination with terms of geometry is already attested in Neo-Latin, as shown by a considerable number of lemmas recorded by Rammingner’s *Neulateinische Wortliste* (e.g., *semicircumferetia* ‘semicircle’)<sup>5</sup> and the *Dictionary of Renaissance Latin* by René Hoven (e.g., *semiglobus* ‘hemisphere’). Interestingly, both dictionaries also record non-specialized terms used in literary prose, in which *semi-* expresses the attenuative value that will also emerge in Italian a few centuries later (e.g., *semierror* ‘half mistake’ or ‘slight mistake’ and *semifractus* ‘half broken’ in Hoven’s dictionary and *semicerto* ‘half certain’ or ‘not really certain’ and *semilutheranus* ‘semi-Lutheran’ in Rammingner’s *Neulateinische Wortliste*).

In the last two subcorpora, the productivity of *semi-* increases, as shown by the higher number of types and the presence of hapaxes. Focusing on the period from 1692 to 1840, *semi-* appears in a limited number of words belonging to scientific terminologies (i.e., *semi-itono* ‘half tone’ in music and *semi-membranoso* ‘a muscle consisting of one-third of membranous tendon’ and *semitendinoso* ‘a muscle that has a long tendon of insertion’ in

<sup>5</sup> The Rammingner’s *Neulateinische Wortliste* covers a period ranging from 1300 to 1700. It is freely available at the following link: <http://nlw.renaessancestudier.org/words/start.htm> (accessed 30 January 2023).

medicine),<sup>6</sup> where it can denote ‘half’ or ‘partially’. Notably, it also appears in a significant number of hapaxes (16) attested in historical/philosophical essays, where it combines with non-specialized nouns and adjectives, as exemplified in (4).

- (4) a. *E di questo semi-amore, o sia*  
 And of this SEMI + love or be.3SG.SUBJ  
*minore odio del tiranno pel popolo [...]*  
 reduced hatred of the tyrant for the people

‘And about this partial love, namely reduced hatred, felt by the tyrant towards the people’. (Vittorio Alfieri, *Della tirannide*, 1777)

- b. *Quasi che un uomo potesse essere semi-innocente*  
 almost that a man can.3SG.SBJV be.INF SEMI + innocento  
*o semi-reo, cioè semi-punibile e semi-assolvibile.*  
 or SEMI + guilty that is SEMI + punishable and SEMI + acquittable

‘Almost as if a man could be partially innocent or partially guilty, that is partially punishable and partially acquittable’. (Cesare Beccaria, *Dei delitti e delle pene*, 1764)

In (4a), the meaning of *semi-amore* is explicitly defined by the author as a low degree of hatred felt by the tyrant toward the people; in this case, *semi-* refers to a reduced degree of love (with a euphemistic nuance). In (4b), the same value is conveyed in four words including *semi-*: the author discusses the possibility of a man being partially guilty and therefore both partially punishable and acquittable (which he finds absurd).

Similarly, in the last subcorpus, *semi-* mostly occurs in historical or political essays where it combines with a still restricted number of adjectives and nouns (i.e., 10 A, 7 N, and 8 N|A), where it refers to a reduced degree of a category (e.g., *semi-indipendenza* ‘partial independence’; [*società*] *semi-organizzata* ‘partially organized society’), as shown in (5).

- (5) *Il popolo che insorge, per avere forti probabilità*  
 the people that rebel.3SG.PRS for have.INF strong possibility  
*di vittoria, deve già godere di una semi-*  
 of victory must already enjoy.INF of a SEMI-

<sup>6</sup> Notably, both *semitonus* and *semimembranosus* are already attested in Neo-Latin (see respectively Hoven’s dictionary and Ramminger’s *Neulateinische Wortliste*).

<b>indipendenza,</b>	<i>in maniera</i>	<i>che</i>	<i>una</i>	<i>parte</i>	<i>di</i>	<i>esso</i>
independence	in way	that	a	part	of	it
<i>sia</i>	<i>militarmente</i>	<i>bene</i>	<i>organizzata.</i>			
be.3SG.SBJV	militarily	well	organized			

‘In order to have a strong chance of victory, the rising people must already enjoy a partial independence, so that a part of them is militarily well organized.’ (Gaetano Mosca, *Elementi di scienza politica*, 1896–1923)

Notably, in two cases ([*libro*] *semi-aperto* ‘partially open’ and *semi-accettazione* ‘partial acceptance’), *semi-* occurs in text genres that are closer to everyday language, namely literary prose and personal letters, as illustrated in (6) with the case of *semi-accettazione*.

- (6) *Un intimo senso di semi-accettazione, come se il*  
a intimate sense of SEMI + acceptance as if the  
*male che impreca fosse in natura ed*  
evil that curse.3SG.PRS be.3SG.SUBJ in nature and  
*inevitabile.*  
unavoidable

‘[...] an intimate sense of partial acceptance, as if the evil he curses were natural and inevitable’. (Massimo D’Azeglio, *Racconti, leggende e ricordi della vita italiana*, 1856)

The example in (6) highlights that the prefix *semi-* is also used in everyday language to express a partial accomplishment of an event conveyed by the base (see Grandi & Iacobini 2008). However, such cases are still very marginal, at least in the written language. The emergence of evaluative values related to approximation can be therefore considered more recent and linked to the use of *semi-* in a broader range of texts/contexts and in combination with a wider variety of bases. A cursory analysis of data extracted from the *la Repubblica* corpus (1985–2000) and Timestamped JSI web corpus (2014–2016) allows to observe in which contexts *semi-* is used and what kind of evaluative values it conveys in everyday language. Starting with the formal properties, Tab. 5 illustrates how the occurrences of *semi-* are distributed with respect to the input and output lexical categories as well as the complexity of the base.



**Tab. 5:** Formal features of words containing *semi*- in Contemporary Italian

	<i>la Repubblica</i> corpus (1985–2000)	Timestamped JSI web corpus (2014–2016)
Input category		
Adjective	63 tokens; 46 types	81 tokens; 54 types
Noun	37 tokens; 27 types	19 tokens; 18 types
Output category		
Adjective	63 tokens; 46 types	81 tokens; 54 types
Noun	37 tokens; 27 types	19 tokens; 18 types
Complexity of the base		
Complex	63 tokens; 45 types	76 tokens; 51 types
Simple	37 tokens; 28 types	24 tokens; 20 types

The data extracted from the *la Repubblica* corpus and those extracted from the more recent Timestamped JSI corpus are comparable and show the same trends. Regarding the input/output lexical categories, in most examples, *semi*- is attached to adjectives (especially relational adjectives, e.g., [*ordine religioso*] *semi-monastico* ‘SEMI + monastic religious order’) to form adjectives and, secondarily, to nouns to form nouns (e.g., *semi-libertà* ‘SEMI + freedom’). Contrary to what Masini & Micheli (2020: 382–383) observed with respect to *simil*-, the addition of *semi*- does not induce a change in the category of the base. As for the complexity of the bases, most bases are morphologically complex words (e.g., *semi-clandestinità* ‘SEMI + secrecy’, where *clandestinità* is a noun derived through the suffix *-ità*).

Turning to the semantics of words containing *semi*-, the analysis of each token from both corpora reveals that *semi*- can express the following values:

1. ‘attenuation, a reduced degree of X-hood’ (i.e., the value attested in the last subcorpus of CODIT, e.g., in *semi-accettazione* ‘partial acceptance’). The speaker wants to emphasize the attenuated degree of the entity/quality/event expressed by the base (sometimes with a derogatory or euphemistic nuance): the *semi*- word still falls within the category expressed by the base but has attenuated properties (e.g., *semi-addormentato* ‘partially asleep’, ‘SEMI + asleep’). In most cases, such as *semi-addormentato*, the attenuation is due to the expressive needs of the speaker and is strictly context-dependent; however, a certain number of words containing *semi*- that

indicate established categories<sup>7</sup> are attested (e.g., [*pistola*] *semi-automatica* ‘semi-automatic pistol’). The function expressed by *semi-* in these cases can be defined as *kin-categorization*, adopting the term proposed by Masini and Micheli (2020: 384) to describe similar uses of *simil-* (i.e., “Y results in a separate (established) entity closely related to X, in that it is defined starting from X and deviating from it”).

2. ‘resemblance/imitation’ (i.e., values belonging to the domain of approximation according to Masini & Micheli 2020):<sup>8</sup> the speaker questions the inclusion of the word containing *semi-* within the category expressed by the base due to the lack of certain fundamental properties (e.g., *semi-relazione* ‘pseudo-relationship’, ‘SEMI + relationship’)

As common in the study of qualitative evaluative values (Grandi 2017), the interpretation of the evaluative constructions is not always easy since it requires understanding the speaker’s purposes from the analysis of the context: some cases have therefore been marked as ‘doubtful’. Tab. 6 sums up the quantitative results of the semantic annotation of each token in the two corpora.

**Tab. 6:** Semantic classification of words containing *semi-*

	<i>la Repubblica</i> corpus (1985–2000)	Timestamped JSI web corpus (2014–2016)
‘attenuation’	58 tokens; 42 types	47 tokens; 41 types
of which ‘kin-categorization’	15 tokens; 9 types	24 tokens; 14 types
‘resemblance/imitation’	24 tokens; 22 types	25 tokens; 21 types
doubtful cases	3 tokens; 3 types	4 tokens; 4 types

As already observed for formal features, the two corpora show a similar distribution of occurrences, suggesting that the functions performed by *semi-* are stable over the time period considered. The more common value conveyed by *semi-* in both corpora is attenuation; as demonstrated by the example in (7a), *semi-* expresses a reduced degree of the category conveyed by the base, which is in this case a darkness that the speaker describes as partial.

<sup>7</sup> By “words referring to established categories”, we mean words referring to stable entities with specific characteristics that are usually part of a taxonomy (e.g., in [*pistola*] *semi-automatica*, *semi-automatica* indicates a specific type of gun with precise characteristics). On the other hand, the form *semi-addormentato* is created by the speaker in context for expressive needs and reflects his subjective perception.

<sup>8</sup> Specifically, ‘resemblance’ is defined as “Y refers to an entity that is merely similar to X without being an X” while ‘imitation’ is defined as “Y refers to an entity that is meant to imitate or reproduce X (without being a genuine X)” (Masini & Micheli 2020: 383).

Whereas in the case of *semi-oscurità* the attenuation is linked to the specific context, *semi-presidenziale* in (7b) indicates a stable category, namely a specific type of government system.

- (7) a. *La semi-oscurità permette di nascondere*  
 The SEMI+darkness permit.3SG.PRS of hide.INF  
*i numerosi occhi lucidi.*  
 the numerous eyes shining.  
 ‘The partial darkness permits hiding the many shining eyes’. (*la Repubblica* corpus)
- b. *Noi per esempio siamo per un sistema*  
 we for example be.3SG.PRS for a system  
*semi-presidenziale alla francese.*  
 SEMI + presidential to the French.  
 ‘For example, we stand for a French-style semi-presidential system’.  
 (*la Repubblica* corpus)

Besides the attenuative function, the prefix *semi-* also expresses values ascribable to the domain of approximation (i.e., ‘resemblance’ and ‘imitation’) when the speaker questions the inclusion of an entity or a quality to the category expressed by the base, as represented in (8).

- (8) a. *Tinder è sia una app sia il relativo*  
 Tinder be.3SG.PRS both a app both the related  
*semi-social network, ma è soprattutto un servizio*  
 SEMI + social media but be.3SG.PRS above all a service  
*di dating online.*  
 of dating online  
 ‘Tinder is both an app and the related near-social media, but it is above all an online dating system’. (Timestamped JSI web corpus)
- b. *Parla, e sotto una cantilena semi-lombarda*  
 speak.3SG.PRS and under a singsong SEMI + Lombard  
*riecheggia un inconfondibile accento meridionale.*  
 echo.3SG.PRS an unmistakable accent southern  
 ‘(She) speaks, and under a Lombard-like singsong echoes an unmistakable southern accent’. (*la Repubblica* corpus)

In both examples, *semi-* expresses the idea that the entity (or the quality) referred to by the speaker lacks certain prototypical properties of the category expressed by the base and is therefore outside it. In (8a), Tinder is defined by the speaker as a ‘near-social media’ because, despite showing resemblance in some respects, it differs from prototypical social media (e.g., Facebook), since it has been primarily designed to facilitate online dating. In (8b), the speaker describes a woman who speaks with an accent that has some traits of the Lombard variety but not to the point of sounding authentic; in fact, the speaker himself still recognizes the typical features of the southern variety. Moreover, it is not clear from the context whether the woman is pretending to speak with a Lombard accent or not; in fact, *semi-* allows both an interpretation in terms of ‘imitation’ (i.e., a Lombard-like accent) and an interpretation in terms of ‘fakeness’ (i.e., a fake-Lombard accent; see the notion of *disproximation* proposed by Cappelle, Daus & Hartmann 2023)<sup>9</sup>.

Interestingly, *semi-* also appears in combination with other morphological elements that express values pertaining to approximation. In the example given in (9), *semi-* combines with *quasi-* (lit. ‘almost’) to reinforce the speaker’s judgement that the militants of the Five Star Movement pretend to have anti-EU positions.

- (9)      *Sappiamo*      *bene*      *che*      *i*      *pentastellati*      *non*      *sono*  
          know.1PL.PRS   well   that   the   5SM supporters   not   be.3PL.PRS
- molto*   *veritieri*              *nelle*   *loro*   *istanze*              ***semi-quasi-antiUe.***  
          very   truthful              in the   their   demands              SEMI + almost + antiEU

‘We know that the Five Star Movement militants are not very truthful in their semi-almost-antiEU demands’. (*la Repubblica* corpus)

To sum up, the cursory analysis carried out on the two corpora of Contemporary Italian revealed that the main function performed by the prefix *semi-* is attenuation, which speakers utilize to signal the reduced degree of a given entity or quality. In such cases, the use of *semi-* does not imply a speaker’s judgement on whether the entity belongs to a particular category, but it is rather a measurement in terms of degree internal to the category conveyed by the base. Besides the attenuative function, *semi-* is used when speakers want to

<sup>9</sup> According to Masini & Micheli (2020: 383), ‘fakeness’ is the value expressed when “Y refers to an entity that is meant to imitate X (without being a genuine X) in order to be taken as an X”.

emphasize that a certain entity or quality resembles or imitates the one expressed by the base but is not precisely ascribable to it. In those cases, *semi*- conveys values pertaining to the domain of approximation, i.e., ‘resemblance’, ‘imitation’, and ‘fakeness’. The two functions (attenuation and approximation) are therefore distinguishable in that, in the former case, the speaker does not question the inclusion of an entity (or a quality) to a category, unlike in the latter case. Needless to say, the two functions are very close and can be seen as situated along a continuum extending from attenuation (when the new concept does not exhibit all the prototypical properties but is still part of the category) to approximation (when the new concept is close to but does not belong to the category expressed by the base, in that it lacks properties that are considered essential).

### 3.2 Analysis of the prefix *para*-

Compared to *semi*-, the prefix *para*- is less frequently attested in CODIT, as shown in Table 7.

**Tab. 7:** Type/token frequencies and the number of hapaxes of words containing *para*- within CODIT

<i>para</i> -	1 (13 <sup>th</sup> c.–1375)	2 (1376–1532)	3 (1533–1691)	4 (1692–1840)	5 (1841–1947)
Types	3	4	6	7	9
Tokens	29	75	264	447	323
Hapaxes	1	2	1	1	2

The type frequency of *para*- is significantly low among all subcorpora: in particular, starting from Old Italian, *para*- occurs in learned loanwords from Latin that were probably semantically and formally opaque already at this stage, e.g., *parabola* ‘parable’ (from Lat. *parabōla*) and *parafrasi* ‘paraphrase’ (from Lat. *paraphrāsis*). Over the following centuries, these words became very frequent in texts, leading to an increase in the token frequency of the prefix. In contrast to what was observed in the case of *semi*-, Neo-Latin dictionaries do not list many examples where *para*- is productively used to form specialized or non-specialized terms; the dictionaries include a number of Ancient Greek loanwords or words created following the Ancient Greek model (e.g., *paraphronesis* ‘delirium’ or ‘madness’, *paragrammatismus* ‘replacing one letter with another’, *paratypum* ‘counterfeit money’, *parachronismus* ‘contradiction in the chronological order’, etc.), where *para*- suggests an anomaly or a deviation with respect to the meaning conveyed by the base.

Only in the last two subcorpora of CODIT one can find words that do not have classical/Neo-Latin origins; these include *parametro* ‘parameter’ (from French *paramètre*), *paraselene* ‘each of the luminous discs that appear on the left and the right of the moon’ (lit. beside the moon), and *paratartarico* ‘paratartaric (acid)’. The last two words belong to scientific terminologies and exemplify two different semantic values conveyed by *para-*. In *paraselene*, *para-* has the original locative value (i.e., ‘proximity’), while in *paratartarico*, which is used to refer to a type of acid that has some (but not all) of the properties of tartaric acid, it expresses ‘similarity’. In both cases, *para-* is used to categorize/identify a given entity based on another entity to which it is close or similar, thus performing a kincategorization function.

The data extracted from CODIT only suggest that *para-* is used in scientific terminologies without giving any information on its usage in non-scientific varieties, which can be considered absent or extremely limited in the period ranging from the 13<sup>th</sup> century to 1947.

Turning to the most recent period, the data extracted from the *la Repubblica* corpus and Timestamped JSI corpus allow us to observe how *para-* is used in present-day Italian. Starting with the analysis of the formal features (see Table 8), *para-* tends to be attached to adjectives to form adjectives. The adjectives used as bases are mostly relational (e.g., [*cellule*] *para-terroristiche* ‘terroristic-like cell’), but instances of qualifying adjectives are still attested (e.g., [*coppie*] *para-famose* ‘pseudo-famous couples’).

**Tab. 8:** Formal features of words including *para-* in Contemporary Italian

	<i>la Repubblica</i> corpus (1985–2000)	Timestamped JSI web corpus (2014–2016)
Input category		
Adjective	86 tokens; 51 types	85 tokens; 58 types
Noun	12 tokens; 12 types	14 tokens; 11 types
Proper Noun	2 tokens; 2 types	1 token; 1 type
Output category		
Adjective	86 tokens; 51 types	86 tokens; 59 types
Noun	14 tokens; 14 types	14 tokens; 11 types
Complexity of the base		
Complex	93 tokens; 56 types	80 tokens; 53 types
Simple	7 tokens; 7 types	20 tokens; 17 types

In most cases, the use of *para-* does not lead to a change of lexical category; however, in the most recent corpus, when combined with a proper noun, *para-* results in an attributive construction, as depicted in (10).

- (10) *I fotogiornalisti che scelgono di imprimere*  
 the photojournalist REL choose.3PL.PRS of impress.INF  
*un' impronta para-Instagram ai propri reportage*  
 a mark PARA-Instagram to the their own reports  
 'Photojournalists who choose to impress an Instagram-style mark on their reports'.  
 (Timestamped JSI web corpus)

Although under very limited conditions, the prefix *para-* seems to have category-changing properties, similar to what was already observed in the case of *simil-* (see Masini & Micheli 2020: 383). As for the complexity of the bases, *para-* clearly tends to combine with morphologically complex bases (especially adjectives, e.g., *para-bancario* 'para-banking' where *bancario* is an adjective derived through the suffix *-ario*).

With regard to semantics, the analysis of each token from both corpora reveals that *para-* can mostly convey the following values:

1. 'subordinate to X, unofficial X': this value appears in more established words in which the base is a noun referring to an institution, a social group, or a professional figure (e.g., [*gruppo*] *para-militare* 'paramilitary, referring to a group that is organized like an army but is not official and often not legal' and *paramedico* 'a person whose training is similar to that of a nurse and who helps to do medical work').
2. 'resembling X': in these cases, *para-* performs an approximating function in that it signals that the entity referred to by the speaker is close to a given category but not fully included in it (e.g., [*intimidazione*] *para-mafiosa* 'mafia-like intimidation'). In some contexts, *para-* may also convey a pejorative nuance (e.g., *para-processo* 'pseudo-trial').

Table 9 summarizes the quantitative results of the semantic annotation of each token in the two corpora.

**Tab. 9:** Semantic classification of words containing *para-*

	<i>la Repubblica</i> corpus (1985–2000)	Timestamped JSI web corpus (2014–2016)
‘subordinate to X’	51 tokens; 20 types	44 tokens; 23 types
‘resembling X’	49 tokens; 44 types	56 tokens; 48 types

The two functions are almost equally distributed among the occurrences analyzed. In particular, the value ‘subordinate to X’ appears in words that are more entrenched within the lexicon (e.g., *para-militare* ‘paramilitary’: 20 occurrences within the *la Repubblica* corpus and 10 occurrences in the Timestamped JSI web corpus). Turning to the approximating function, it is mostly attested in nonce words where it fulfills expressive needs related to the specific context; in such cases, the speaker wants to refer to something that resembles a certain category, as illustrated in (11) where *para-leopardiana* indicates a philosophical conclusion that is in the style of the Italian poet Giacomo Leopardi.

- (11) *Squisitamente para-leopardiana la conclusione filosofica [...]*  
 Purely PARA-Leopardian the conclusion philosophical  
 ‘(It’s) a typically Leopardian-like philosophical conclusion [...]’ (*la Repubblica* corpus)

Although it is not easy to identify from the context, in some cases *para-* conveys a pejorative nuance, as shown in (12) where the speaker criticizes people who think they are health-conscious but actually are not.

- (12) *Il resto sono balle, anche quelle dei para-salutisti*  
 the rest be.3PL.PRS bullshit also those of the PARA-health-conscious  
*che ogni sei mesi scoprono che mangiare*  
 rel every six months find\_out.3PL.PRS that eat.inf  
*dietetico fa bene al cuore*  
 dietetic make.3SG.PRS good to the heart  
 ‘The rest is bullshit, even those of the pseudo-health-conscious who, every six months, find out that dietary eating is good for the heart’. (Timestamped JSI web corpus)

Finally, as already noted for *semi-*, *para-* may combine with other prefixes or prefixoids expressing values pertaining to approximation in order to reinforce the evaluative meaning and make the utterance vaguer (cf., ‘intentional vagueness’ in Voghera 2012), as shown in (13).



- (13) *Uno spettacolo di cui erano fino a un' ora*  
 a show of REL be.3PL.PST until to a hour  
*fa protagonisti (o quasi-, pseudo- para-protagonisti)*  
 ago protagonists or almost pseudo PARA-protagonists  
 'A show in which they were protagonists (or almost, pseudo-, para-protagonists) until an hour ago.' (*la Repubblica* corpus)

To sum up, this cursory analysis has shown that the prefix *para*- is productively used in everyday Italian; notably, it appears with a similar frequency both in words referring to established categories with the meaning 'subordinate to X' and in nonce words where it conveys the approximative value 'resembling X'. The relationship between the two values will be discussed in the following section.

#### 4. Discussion and conclusions

This study has enriched the literature on the development of approximating values in prefixes of classical origins. By analyzing the morphological and semantic properties of words containing *semi*- and *para*- in both diachronic and synchronic corpora, we found that both prefixes originally pertain to specialized terminology where they perform the categorizing function of creating new terms in scientific taxonomies. Focusing on *semi*-, only from the 17<sup>th</sup> century onwards, it is also attested in historical/political essays and (more rarely) in literary prose and personal letters, where it expresses attenuation by referring to a reduced degree of a category. Both values (scientific categorization and attenuation) are already attested in Renaissance Neo-Latin, which, as other studies have already shown (see, e.g., Rainer 2008 on *neo*-), anticipates some of the semantic values with which prefixes of classical origin are used in Italian in the later centuries. The synchronic analysis based on two Contemporary Italian corpora has shown that the main function of *semi*- is attenuation, which is performed both in words attenuated by the speaker for subjective evaluations and in words denoting established categories of reduced degree compared to those expressed by the base. This function does not imply a real qualitative evaluation by the speaker, but it is rather a measurement in terms of degree, closer to what Grandi (2002) calls 'objective evaluation'. Interestingly, besides the attenuative function, *semi*- is also used in contexts where the speaker wants to emphasize that a certain entity/property/event exhibits some

features of a certain category but not enough to be part of it. In such cases, the prefix performs approximative functions such as ‘resemblance’ and ‘imitation’, in that it refers to entities that are close to a category but not part of it. Indeed, the analysis of the contexts of use of *semi-* suggests that attenuation and approximation can be regarded as two distinct domains lying along a continuum, ranging from the former, where the entity mentioned by the speaker is inside the category but is downgraded, to the latter, where the entity is close but outside the category.

Turning to the prefix *para-*, the data extracted from CODIT are very scarce and seem to indicate a limited use of the prefix during the earliest centuries of the history of Italian. In the oldest subcorpora, it is attested only in opaque learned loanwords from Ancient Greek, and even Neo-Latin dictionaries record mostly words of Ancient Greek origin or neoclassical compounds, where the prefix indicates an anomaly or a deviation with respect to the meaning conveyed by the base (e.g., *parachronismus* ‘contradiction in the chronological order’). This value is presumably related to the original semantic functions performed by *para-* in Ancient Greek alongside the locative one (i.e., ‘contrary to’ and ‘other than’). Within CODIT, the only attestations of *para-* being used productively are in the field of specialized terminologies, where it may convey the original locative meaning (e.g., *parase-lene*, lit. ‘beside the moon’) or indicate similarity (e.g., *paratartarico* ‘tartaric-like acid’). While *para-* is productively used only in specialized terminology within CODIT, the analysis of Contemporary Italian data indicates that it appears in everyday language both in frequent and well-established words, which refer to something or someone that is subordinate to another entity conveyed by the base (such as *para-militare* ‘paramilitary’), and in nonce words where it serves an approximating function (i.e., the value ‘resembling’). In both cases, *para-* indicates that an entity has some properties of a category exemplar but not all of them (or at least not the most prototypical ones), so it may be similar but of a lower level or just resemble it.

This investigation has shown that the original values of the two prefixes play a role in the development of specific semantic nuances related to evaluation and particularly approximation. In the case of the prefix *semi-*, its original value ‘degree measurement’ resulted in an attenuative function and, in some contexts, in approximative values related to resemblance and imitation, on the basis that what has attenuated properties resembles the

full entity or quality. As for *para*-, the original locative value has left its traces in the current meaning ‘subordinate to X’ since subordination presupposes a closeness between the two entities; moreover, what is subordinate to X resembles X but does not have all its qualities. Finally, this study confirms that, alongside similarity items (see Masini, Micheli & Huang 2018 for an overview), fake items (see Van Goethem, Norde & Masini 2021; Van Goethem & Norde 2020), and diminutives (see, e.g., Stosic & Amiot 2023), spatial proximity items and degree items are possible sources of approximating morphological markers.

To conclude, this paper has highlighted some relevant issues that are worthy of further investigation. First, this study focused on the semantics of the prefixes *para*- and *semi*-, while very little has been said about their (possible) pragmatic effects: further research should verify whether (and to what extent) the prefixes *para*- and *semi*- can be used by speakers to modify the illocutionary force of speech acts, following the morphopragmatic perspective outlined by Dressler & Merlini Barbaresi (1994) (see, e.g., Mattiello 2009 on the mitigating function performed by negative prefixes). Second, a formal issue needs to be deepened: this study confirms a preference for prefixation in the expression of approximation in Italian, as shown by the great number of approximating prefixes or prefixoids identified by Masini and Micheli (2020: 393–397; i.e., *finto*-, *para*-, *semi*-, *simil*-, *quasi*-, *non*-, *mezzo*-) in contrast to a smaller number of suffixes conveying this function (i.e., *-oide*-, *-esco*-, *-eggianti*-, see Micheli 2022b). This issue deserves further study with a deeper focus on morpho-phonological properties shown by the bases. Preliminarily, we merely note that suffixation shows more constraints in terms of combinability with morphological complex bases, while prefixes exhibit a greater autonomy from the bases and can therefore (presumably) represent the favored strategy for constructions created in the moment by speakers. Finally, the adoption of a cross-linguistic (particularly Europe-wide) perspective seems crucial in the study of neoclassical elements in order to verify whether the emergence of evaluative meanings is attested in other European languages.

## References

- Baroni, Marco, Silvia Bernardini, Federica Comastri, Lorenzo Piccioni, Alessandra Volpi, Guy Aston & Marco Mazzoleni. 2004. Introducing the “la Repubblica” corpus: A large, annotated, TEI(XML)-compliant corpus of newspaper Italian. *Proceedings of LREC 2004*. <http://www.lrec-conf.org/proceedings/lrec2004/pdf/247.pdf> (accessed 7 June 2022).
- Bortone, Pietro. 2010. *Greek Prepositions: From Antiquity to the Present*. Oxford: Oxford University Press.
- Cappelle, Bert, Robert Daugs & Stefan Hartmann. 2023. The English privative prefixes *near-*, *pseudo-*, *quasi-* and *sub-*: Approximation and ‘disproximation’. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1). 52–75.
- D’Achille, Paolo. 1990. *Sintassi del parlato e tradizione scritta della lingua italiana*. Roma: Bonacci.
- Dressler, Wolfgang U. & Lavinia Merlini Barbaresi. 1994. *Morphopragmatics. Diminutives and Intensifiers in Italian, German, and Other Languages*. Berlin: De Gruyter.
- Efthymiou, Angeliki, Georgia Fragaki & Angelos Markos. 2015a. Exploring the meaning and productivity of a polysemous prefix: The case of the Modern Greek prepositional prefix *para-*. *Acta Linguistica Hungarica* 62(4). 447–476.
- Efthymiou, Angeliki, Georgia Fragaki & Angelos Markos. 2015b. Exploring the polysemy of the Modern Greek prefix *iper-*. *Morphology* 25(4). 411–438.
- Grandi, Nicola. 2002. *Morfologie in contatto. Le costruzioni valutative nelle lingue del Mediterraneo*. Milano: FrancoAngeli.
- Grandi, Nicola. 2017. Evaluatives in Morphology. In *Oxford Research Encyclopedia of Linguistics*. DOI: [10.1093/acrefore/9780199384655.013.250](https://doi.org/10.1093/acrefore/9780199384655.013.250).
- Grandi, Nicola & Claudio Iacobini. 2008. L’affissazione valutativa nei verbi dell’italiano. In Emanuela Cresti (ed.), *Nuove prospettive nello studio del lessico. Atti del IX Congresso Internazionale SILFI (Firenze, 15-17 giugno 2006)*, 483–490. Firenze: Firenze University Press.
- Grandi, Nicola & Livia Körtevelyessy. 2015. *The Edinburgh Handbook of Evaluative Morphology*. Edinburgh: Edinburgh University Press.
- Iacobini, Claudio. 2004. Composizione con elementi neoclassici. In Maria Grossmann & Franz Rainer (eds.), *La formazione delle parole in italiano*, 69–95. Tübingen: Niemeyer.
- Iacobini, Claudio, Aurelio De Rosa & Giovanna Schirato. 2017. Criteri e strategie di classificazione morfosintattica dei testi del corpus MIDIA. In Paolo D’Achille & Maria Grossmann (eds.), *Per la storia della formazione delle parole in italiano. Un nuovo corpus in rete (MIDIA) e nuove prospettive di studio*, 33–55. Firenze: Cesati Editore.
- Luraghi, Silvia. 2003. *On the Meaning of Prepositions and Cases: The expression of Semantic Roles in Ancient Greek*. Amsterdam/Philadelphia: John Benjamins.
- Masini, Francesca & M. Silvia Micheli. 2020. The morphological expression of approximation. The emerging *simil-* construction in Italian. *Word Structure* 13(3). 396–427. DOI: [10.3366/word.2020.0176](https://doi.org/10.3366/word.2020.0176).
- Masini, Francesca, M. Silvia Micheli & Shanshan Huang. 2018. Attenuation and categorization: A cross-linguistic study of similarity items. Paper presented at the Workshop “Attenuated qualities in a cross-linguistic perspective”, 51st Annual Meeting of the Societas Linguistica Europaea (SLE 51), Tallinn, 29 August–1 September.

- Mattiello, Elisa. 2009. A morphopragmatic analysis of English and Italian negative prefixes. *Studi e Saggi Linguistici* XLVII. 125–156.
- Micheli, M. Silvia (2022a). CODIT. A new resource for the study of Italian from a diachronic perspective. Design and applications in the morphological field. *Corpus* 23. *Special Issue “Corpus et Données en Morphologie”*. DOI: [10.4000/corpus.7306](https://doi.org/10.4000/corpus.7306).
- Micheli, M. Silvia (2022b). Italian approximative suffixes in diachrony. Poster presented at the 13<sup>th</sup> Mediterranean Morphology Meeting. Rhodes, 19–21 May.
- Rainer, Franz. 2008. Neo- and neo-Latin. *Word Structure* 1(1). 53–64.  
DOI: [10.3366/E1750124508000068](https://doi.org/10.3366/E1750124508000068).
- Rainer, Franz. 2015. Intensification. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-formation. An International Handbook of the Languages of Europe*, vol. 2, 1339–1351. Berlin: De Gruyter.
- Stosic, Dejan & Dany Amiot. 2023. Why does morphological augmentation not lead to approximation? The case of denominal evaluative nouns. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1). 177–203.
- Van Goethem, Kristel & Muriel Norde. 2020. Extravagant “fake” morphemes in Dutch. Morphological productivity, semantic profiles and categorical flexibility. *Corpus Linguistics and Linguistic Theory* 16(3). 425–458.
- Van Goethem, Kristel, Muriel Norde & Francesca Masini. 2021. The fate of ‘pseudo-’ words: A contrastive corpus-based analysis. Paper presented at UCCTS 2021 (Using Corpora in Contrastive and Translation Studies), Bertinoro, 9–11 September 2021.

M. Silvia Micheli  
Università degli studi di Milano – Bicocca  
Piazza dell’Ateneo Nuovo 1  
20126 Milano, Italy  
[maria.micheli@unimib.it](mailto:maria.micheli@unimib.it)



This is an open access publication. This work is licensed under a Creative Commons Attribution CC-BY 4.0 license. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>