InTens – A dataset of Italian intensified derivatives. Description and application in a productivity study

Ivan Lacić
University of Bologna
ivan.lacic2@unibo.it

Abstract

The paper introduces *InTens*, a dataset of Italian intensified adjectival derivatives formed with six evaluative prefixes, namely *arci*, *extra*, *iper*, *stra*, *super*, and *ultra*. Initially, we delineate the process of data extraction and filtration. Subsequently, we address the polyfunctional characteristics of the evaluative prefixes, with semantic annotation of derivatives into two macro-categories: INTENSIFICATION and NON-INTENSIFICATION. After discussing the annotation results, the application of *InTens* is demonstrated through an investigation of the morphological productivity of the prefixes. The analysis underscores the variability in productivity contingent upon the semantic function of the prefixes, an aspect most often overlooked in productivity research.

1 Motivation

While large-scale resources providing a comprehensive view of the word-formation system are available for languages such as French (e.g., *Démonette-2* (Namer et al., 2023)), the situation for Italian, particularly concerning derivation, remains comparatively underdeveloped (for instance, Morph-It! (Zanchetta and Baroni, 2005) does not encompass derivational structure and derivationally related lemmas). This study seeks to make a contribution toward addressing this gap by developing a small-scale resource providing a coherent set of Italian adjectival derivatives expressing INTENSIFICATION, formed with six evaluative prefixes: *arci*, *extra*, *iper*, *stra*, *super*, and *ultra*.[1] This dataset is believed to be novel not only for Italian, but conceptually also for more extensively studied languages, as the majority of intensified derivatives do not form entrenched words with stable references in the lexicon and are therefore frequently omitted from existing resources.

The utility of *InTens* becomes particularly apparent when analyzing the six prefixes as formally suppletive morphemes that produce near-synonyms. In the context of Italian, the limited existing literature on evaluative prefixes has focused exclusively on cataloging these elements (Montermini, 2008), examining them as isolated phenomena (Napoli, 2012), or through a contrastive lens (Calpestrati, 2017), without addressing their competitive interactions. At its most basic, competition "refers to the fact that speakers routinely have to make a choice between alternative ways of realizing a certain concept" (Gardani et al., 2019, p. 4). A particular case of morphological competition is known as affix rivalry. Affix rivalry refers to the interaction between two or more affixes that, in at least some of their uses, can generate words of identical or similar semantic types (Guzmán Naranjo and Bonami, 2023; Huyghe and Varvara, 2023; Nagano et al., 2024). Evaluative morphology (EM) emerges as a particularly apt domain for the investigation of affix rivalry, as a single evaluative function is often expressed by multiple formal exponents,

[1] Our primary focus lies on evaluative prefixes that convey qualitative evaluation (+GOOD). Nevertheless, due to the potential difficulty in discerning between qualitative and quantitative evaluation (Iacobini, 2004; Napoli, 2012), we also considered prefixes capable of expressing both +GOOD and +BIG. To identify such prefixes, a thorough investigation of the existing literature was undertaken. Although space constraints preclude detailing the complete rationale behind selecting six specific prefixes, it should be noted that after the application of numerous exclusion criteria (e.g., the exclusion of predominantly quantitative prefixes such as *maxi-* and *mega-*, and the exclusion of prefixes primarily conveying spatial meaning such as *sopra-* and *sur-*) and restricting our selection to prefixes that productively combine with adjectives ţtepiacobini2004, we were left with the six prefixes under examination.

Proceedings of the 5th International Workshop on Resources and Tools for Derivational Morphology (DeriMo 2025),
pages 73–85, Fribourg, Switzerland, 4-5 September 2025.

thus violating the blocking principle (Grandi, 2015). Many morphological systems, especially those of Standard Average European languages, abound with synonymous evaluative affixes that can be appended to the same base, resulting in derivatives that are semantically largely comparable (Körtvélyessy, 2015). While the examined six prefixes exhibit extensive polysemy, deriving words of various semantic functions – SPATIAL LOCATIVE (e.g., *extraurbano* 'extraurban'), NON-SPATIAL LOCATIVE (e.g., *stragiudiziale* 'extrajudicial'), HIERARCHY (e.g., *arcivescovile* 'archbishop's'), EXCESS (e.g., *strapagato* 'overpaid'), INTENSIFICATION (e.g., *ultramoderno* 'ultra-modern'), etc. – INTENSIFICATION seems to be the only domain in which all six prefixes compete. This is not surprising, as predominantly pragmatic communicative objectives that are met through INTENSIFICATION (differently from prototypical derivational morphology) promote the emergence of morphological competition as an acceptable pleonastic feature of the system (Dressler et al., 2019; Merlini Barbaresi and Dressler, 2020).

Given that rival affixes often exhibit differences in productivity (Gaeta and Ricca, 2015), productivity may serve as one of the distinguishing factors in rivalry. Therefore, it is crucial to quantify it using a valid dataset, namely one that comprises derivatives exclusively conveying INTENSIFICATION. Although it has long been recognized that the focus of productivity studies should be directed towards the productivity of semantic functions within a morphological pattern, rather than the pattern as a whole (Kastovsky, 1986), this aspect is frequently overlooked. Consequently, investigations into productivity that incorporate considerations of affix polysemy are more the exception than the rule (for a polysemy-aware approach within the field of EM, see, e.g., Efthymiou et al., 2015).

2 Dataset creation

To understand the properties of intensified derivatives, extensive data annotated for instances of INTENSIFICATION are essential. Section 2.1 outlines the methodology employed in the construction of this dataset, whereas the annotation procedure is presented in Section 2.2.

2.1 Data extraction

The presented dataset is based on the `itWaC` corpus (Baroni et al., 2009), comprising approximately 2 billion words. While not the largest or most recent web corpus, `itWaC` remains the largest copyright-free, freely downloadable corpus of Italian.[2]

As a first step in the dataset creation, adjectives formed with one of the six specific strings – *arci*, *extra*, *iper*, *super*, *stra*, and *ultra* – were extracted. For each of the six strings under exam, constructions in three orthographic variants, namely solid spelling (e.g., *ipercostoso* 'superexpensive'), hyphenation (e.g., *iper-costoso*), and open spelling (e.g., *iper costoso*) were taken into consideration. Instances in which an evaluative prefix follows the base, assuming an attributive use, were not considered. Although such a construction is possible with nominal bases (e.g., *il finalone iper* 'hyper grand finale'), it is typically not anticipated with adjectival bases, as in Italian adjectives generally do not modify other adjectives (e.g., **una casa costosa super* 'a super expensive house') (Montermini, 2008).

Due to the extraction regime not considering the nature of the extracted word – whether a real derivative or just a unit starting with one of the strings formally identical with affixes – cleaning was required. As an initial refinement step, we excluded adjectives that begin with one of the six target strings, where such a string serves as a component of the base rather than functioning as a prefix (e.g., *stra* in the adjective *stradale* 'street$_{\text{REL}}$'). To do this, we cross-referenced extracted occurrences against a list of Italian adjectives obtained from *lo Zingarelli 2024* dictionary (Zingarelli et al., 2023), then removed any matches. Moreover, many of the extracted adjectival "bases" were partial constituents of real words, such as the string *tegico*, a constituent of the adjective *strategico* 'strategic'. To address this, we utilized

[2]While INTENSIFICATION is often associated with spoken language due to its informal characteristics (Ito and Tagliamonte, 2003), a spoken corpus proved unsuitable for our analysis given the limited size of available corpora and the resulting low frequency of relevant derivatives. For instance, a preliminary analysis of the `KIParla` corpus (Mauri et al., 2019) identified only about 60 intensifying prefix occurrences, rendering quantitative analysis unfeasible. That said, a critical distinction between the diamesic and diaphasic dimensions of language use is necessary. We believe that the productivity of our prefixes depends more on the language formality than on the medium itself. The `itWaC` corpus, covering a wide range of text types, is largely independent of this diaphasic variable, making it an appropriate resource.

Wiktionary data and cross-referenced the extracted forms with a list of Italian adjectives from the machine-readable dictionary `kaikki` (Ylonen, 2022). Adjectives absent in the `kaikki` dictionary were excluded, resulting in a more refined list of adjectives. Furthermore, to minimize noise stemming from spurious forms, bases with a frequency less than 5 in the corpus were excluded from the analysis.[3]

Another challenge encountered was the presence of derivatives in which prefixation is anterior to denominal suffixation (e.g., *arcidiocesano* 'archdiocesan' < *arcidiocesi* 'archdiocese'). The standard choice in these situations has been to select only tokens in which the affix is appended as the final element, thereby excluding inner derivations (Fradin et al., 2008; Baayen, 2009). Consequently, it was decided to omit these instances from the analysis.[4] This decision is based on two considerations: first, it is posited that the presence of an already prefixed base in the formation of a derivative impacts the new derivative by analogy; second, this study seeks to examine the productivity of prefixes with exclusively adjectival bases.

Lastly, since a single morphological family may encompass closely related nouns and adjectives, it is often ambiguous which lexeme should be considered the derivative's base (e.g., *supertechnologico* 'supertechnological' could be derived either via suffixation from the noun *supertechnologia* 'supertechnology' or via prefixation from the adjective *technologico* 'technological'). As noted by Bonami and Thuilier (2019), there is frequently no operational method for determining the base of the derivative when the derivational family presents multiple potential "solutions". Therefore, for practical purposes, derivatives that could theoretically be derived through prefixation (as one among the possible mechanisms of word-formation) were not excluded from the analysis. After these refinement steps, and once the data set had been reduced to a manageable size, a final manual verification of derivatives in context, as attested in the corpus, was conducted.

Following the described refinement processes, a final list of prefixed derivatives was obtained. The complete dataset consists of 4,599 derivative types formed with 2,683 adjectival base types. The distribution of the derivatives across prefixes is illustrated in Table 1.

	Tokens	Types	Hapax legomena
arci	1,318	117	81
extra	75,109	722	235
iper	9,695	988	430
stra	20,924	342	163
super	12,888	1,327	528
ultra	19,279	1,103	492

Table 1: Distribution of the derivatives across prefixes.

A notable disparity in the frequency distribution of both tokens and types among prefixes can be observed. Notably, the prefix *extra* is the most frequent prefix by token count at 75,109 occurrences, followed by *stra* with 20,924 occurrences and *ultra* with 19,279 occurrences. Conversely, the prefixes *super* and *iper* register moderate frequencies of 12,888 and 9,695 tokens respectively, whereas the prefix *arci* demonstrates the lowest frequency with 1,318 tokens. Regarding unique types, the prefix *super* possesses the highest number of distinct types, totaling 1,327, while the prefix *arci* contains the fewest with only 117 unique types.

Whilst this dataset can serve as a basis for the examination of derivatives across all semantic values, the specific focus of this study, that is the analysis of competition within the INTENSIFICATION domain, necessitates a focused dataset comprising derivatives that specifically express INTENSIFICATION.

[3]Note that this procedure did not exclude any hapax legomena that might appear with the prefixes as long as the general frequency of the adjectival base in the corpus is ≥ 5.

[4]While the inclusion of inner cycle derivatives may hold psycholinguistic relevance (Plag, 1999), Gaeta and Ricca (2006) observe that when productivity calculations are performed on a fixed token count, as applied in the present study, nearly identical results are obtained irrespective of the inclusion of inner derivations.

2.2 Annotation for semantic function

Given the high specificity of the required annotations, the manual annotation of the derivatives was selected as the preferred methodological approach. While a detailed semantic annotation of the derivatives could potentially reveal interesting insights regarding the competition between prefixes, it was presumed to be extremely labor intensive and, most importantly, not central to this study, which primarily focuses on the phenomenon of rivalry within the context of INTENSIFICATION. For this reason, it was decided to categorize the derivatives into two distinct semantic macro-categories: (i) INTENSIFICATION and (ii) NON-INTENSIFICATION. Annotators had three labels at their disposal: (i) INTENSIFICATION, (ii) NON-INTENSIFICATION, and (iii) TERMINOLOGY. The label TERMINOLOGY refers to the derivatives that are considered part of the scientific lexicon, for which the annotators were uncertain of a semantic value to assign. Furthermore, in instances of ambiguous interpretations (e.g., *iperattivo* 'hyperactive' can be understood both as conveying INTENSIFICATION and EXCESS), annotators were advised to classify the derivative as an instance of INTENSIFICATION, provided it permits such a potential interpretation. This decision stemmed from the understanding that distinguishing between these two interpretations with precision is both challenging and potentially counterproductive given the relatively low frequency of the derivatives being analyzed.

Considering the size of the dataset ($139,213$ tokens), it was split into two parts. The annotations were carried out by the author in conjunction with four co-annotators. The co-annotators are native speakers of Italian, all of whom are either in the process of obtaining or have already obtained a doctoral degree in linguistics. The relatively homogeneous background of the annotators, coupled with their prior linguistic training, is considered advantageous (Artstein, 2017). Each part of the dataset was annotated by three annotators: the author, along with annotators 1 and 2 for the first part, and the author, along with annotators 3 and 4 for the second part. While explicit annotation guidelines in the form of a distinct document were not formulated, comprehensive procedural instructions were provided to each annotator, with consistent communication maintained throughout the annotation process. All issues and inquiries were addressed and resolved collaboratively.

Due to the large size of the initial dataset, annotating individual tokens was not feasible. As a result, the annotations were performed on a type-based level, a methodology we recognize introduces a certain level of approximation.[5] To assess the trustworthiness of the annotations, a qualitative analysis of the annotations was initially conducted, followed by the computation of inter-annotator agreement metrics. Two distinct measures of agreement were initially selected: raw agreement (RA) (Goodman and Kruskal, 1959) and Fleiss' κ (κ_F) (Fleiss, 1981). However, observing the results obtained, a great discrepancy between RA and κ_F values was seen, and it was noted that κ_F encountered a challenge associated with the degree of homogeneity in the annotations, known as the Kappa paradox. In highly homogeneous annotations, where the majority of annotation points fall into a single category (a very common occurrence in our dataset), the marginal probabilities are expected to be very high. Consequently, the expected agreement by chance alone (as measured by κ_F) is also likely to be considerably elevated, and this will, in turn, make κ_F rather low.

To gain a more accurate understanding of the level of agreement among the annotators, an additional metric was introduced, namely Gwet's AC1 (Gwet, 2008). Gwet's AC1 addresses certain limitations

[5] We recognize that type-based annotations should not preclude polysemy, given that certain derivatives may present both intensive and non-intensive interpretations. To (at least partially) address the assumption of monosemy adopted in this study, a random sample of 20 derivative types with a frequency ≥ 10 per prefix was extracted and an analysis on a randomly selected sample of 300 tokens was conducted. Although no instances of genuine polysemous interpretations, encompassing both intensified and non-intensified readings, were identified, two examples classified as TERMINOLOGY allowing also for a non-terminological intensified interpretation were discovered. Specifically, the derivative *iperelastico* appeared both as an instance of INTENSIFICATION meaning 'very elastic' and in a technical context as a term in continuum mechanics. Additionally, the derivative *ultraperiferico* 'ultra-peripheral', was observed in both the terminological reading, referring to a territory of the EU located outside the European continent, and a genuinely intensified reading describing a property for sale in a remote location. As previously noted, both derivatives were adjudicated as TERMINOLOGY by the annotators and subsequently excluded from the analysis, thus not considering the intensifying reading of the two lexemes. It is evident, however, that for more robust conclusions, a significantly larger token sample should be assessed, and we intend to revisit this matter in future research. In the meantime, considering the almost perfect monosemy detected in our $1,800$ token sample, we believe that it is safe to proceed with the analysis.

observed for κ_F, particularly in contexts of high agreement, where κ_F value is typically diminished disproportionately.[6]

	Raw agreement (%)	Fleiss' κ	Gwet's AC1 (95% CI)
arci	98.4	−0.01	0.99 (0.97–1.00)
extra	85.3	0.60	0.86 (0.83–0.89)
iper	87.5	0.32	0.86 (0.83–0.88)
stra	99.4	0.42	0.98 (0.97–1.00)
super	91.3	0.33	0.90 (0.88–0.92)
ultra	89.8	0.60	0.90 (0.88–0.93)

Table 2: Inter-annotator agreement – first part of the dataset.

	Raw agreement (%)	Fleiss' κ	Gwet's AC1 (95% CI)
arci	98.1	0.49	0.99 (0.94–1.00)
extra	80.7	0.60	0.86 (0.82–0.89)
iper	83.9	0.32	0.88 (0.86–0.91)
stra	94.9	0.54	0.96 (0.93–0.99)
super	86.4	0.35	0.90 (0.88–0.92)
ultra	82.5	0.52	0.87 (0.85–0.90)

Table 3: Inter-annotator agreement – second part of the dataset.

Tables 2 and 3 present the values of the three inter-annotator agreement measures applied to the two parts of the dataset. The highest RA and AC1 values were observed for the prefixes *arci* (RA: 98.4/98.1; AC1: 0.99) and *stra* (RA: 99.4/94.9; AC1: 0.98/0.96), while the remaining prefixes also showed consistently high inter-annotator agreement, with all values exceeding the commonly accepted 0.80 threshold for reliable annotations[7] (Brezina, 2018, p. 89). The particularly high levels of agreement for *arci* and *stra* suggest that derivatives formed with these prefixes seem comparatively less polyfunctional, which likely facilitated annotation and contributed to stronger agreement.

The negative κ_F value for *arci* in the first part of the dataset, along with relatively low κ_F values for *iper* and *super*, stand in stark contrast to their high RA and AC1 values. This discrepancy underscores the significant impact that the selection of agreement metrics can have on the interpretation of inter-annotator reliability, emphasizing the necessity of employing multiple agreement measures for acquiring a comprehensive understanding of reliability. The robustness of the findings is substantiated by a correlation analysis between the three agreement measures. This revealed a perfect positive correlation ($\rho = 1$) between RA and AC1 in the second part of the dataset, and a very strong correlation ($\rho = 0.83$) in the first. In contrast, Fleiss' κ displayed a negative correlation with both RA and AC1 across both datasets. This is consistent with the earlier observation, where it was indicated that κ_F tends to underestimate agreement due to the homogeneous nature of the annotations. In conclusion, the consistently high RA ($> 80\%$) and Gwet's AC1 values (≥ 0.86) provide strong evidence of substantial agreement among annotators. It is therefore reasonable to conclude that the annotations are reliable.

The subsequent phase in the development of the dataset involved the identification of intensified derivatives. This process adhered to specific criteria: if two of the three annotators assigned the same tag to a derivative, it was classified accordingly. Conversely, if each annotator assigned a distinct tag to the same derivative, it was categorized as UNCLEAR. Table 4 presents the annotation results for both parts of the dataset.

Out of $4,599$ derivative types, $3,686$ (80.1%) were classified as instances of INTENSIFICATION (e.g., *arcicostoso* 'super-expensive', *strafigo* 'super-cool'), while 809 (17.6%) derivatives were labeled as NON-

[6] While Silveira and Siqueira (2023) champion Gwet's AC1 as the preferred measure for inter-rater agreement in contingency tables, Vach and Gerke (2023) argue that it shouldn't be considered a comprehensive κ-based metric. To sidestep potential methodological debates arising from favoring just one approach, we do not propose AC1 as a replacement for κ-based measures here. Instead, it acts as a complementary metric that offers increased stability in the face of category prevalence effects.

[7] In this study, RA was determined by a strict criterion: only instances where all three annotators assigned the identical label counted as agreement. While this conservative approach captured full unanimity, it inherently excluded cases of partial agreement (e.g., two out of three annotators agreeing). In contrast, AC1 estimates agreement by calculating pairwise consistency across all annotator combinations, then adjusting for chance. Because AC1 does not demand full unanimity, it is sensitive to high levels of partial agreement. Consequently, the observed agreement contributing to AC1 can sometimes surpass the strictly calculated RA, especially when the consensus is strong but not absolute. Furthermore, AC1's chance agreement correction is less affected by category imbalance than traditional κ-based methods. In datasets dominated by one label, this yields more stable, and often higher, adjusted agreement values. Therefore, instances where AC1 slightly exceeds RA should not be interpreted as inconsistencies. Instead, they demonstrate AC1's ability to offer a more inclusive and robust estimate of annotator consensus, particularly in scenarios characterized by high RA and skewed category distributions.

	Types	INTENSIFICATION	NON-INTENSIFICATION	TERMINOLOGY	UNCLEAR
arci	117	116	1	0	0
extra	722	122	592	1	7
iper	988	904	37	31	16
stra	342	336	6	0	0
super	1,327	1,244	51	10	22
ultra	1,103	964	122	3	14
SUM	4,599	3,686	809	45	59

Table 4: Distribution of derivative types across annotation categories.

INTENSIFICATION (e.g., *extraregionale* 'extra-regional', *ultraindividuale* 'extra-individual). Additionally, 45 instances were labeled as TERMINOLOGY (e.g., *ipercontratto* 'over-contracted', *ultralineare* 'ultra-linear'), and 59 instances as UNCLEAR. Given that the TERMINOLOGY and UNCLEAR classes collectively represented only 2.3% of the dataset, and in order to avoid any reliance on subjective authorial judgment concerning their status, they were excluded from further analysis. Finally, a consolidated table was created, containing 3,686 intensified derivative tokens. Table 5 presents the distribution of intensified derivatives per prefix, along with the percentage of intensified tokens and types within the total sample, as illustrated in Table 1.

	Intensified tokens	% of total tokens	Intensified types	% of total types
arci	1,297	98.4	116	99.2
extra	837	1.1	122	16.9
iper	7,930	81.8	904	91.5
stra	18,581	88.8	336	98.3
super	11,167	86.7	1,244	93.8
ultra	8,257	42.8	964	87.4

Table 5: Token and type counts of intensified derivatives.

For certain prefixes, token and type counts change minimally. Specifically, *arci*'s tokens decrease by just 1.6% and its types by 0.8%. Similarly, *stra*, *super*, and *iper* undergo token reductions ranging from 11.2% to 18.2%, and type reductions spanning 1.7% to 8.5%, highlighting their primary role as intensifying prefixes. In contrast, *ultra* and *extra* demonstrate more pronounced reductions when confined to intensifying forms, as evidenced by a 57.2% decrease in *ultra*'s token count, and a sharp 98.9% reduction in *extra*'s, thereby underscoring *extra*'s predominant non-intensifying use. While *extra*'s token and type count reductions are largely proportional, *ultra* displays a unique pattern characterized by a significantly smaller decrease in the type count, measured at 12.6%. This indicates that limiting *ultra* to the INTENSIFICATION domain removes a limited quantity of high-frequency (mainly terminological) types, while largely preserving the diverse, less frequent intensified types, resulting in a comparatively minor change to its overall type count.

The results highlight the importance of annotating for semantic function, as it provides a more nuanced understanding of prefix usage, particularly within the framework of prefix polyfunctionality and rivalry. Without proper semantic annotation, these distinctions would be overlooked, potentially skewing the interpretation of other analyses. Moreover, existing descriptions regarding the usage of intensifying prefixes are challenged. For instance, Calpestrati (2017) suggests that *extra* in Italian primarily functions as an intensifier (regardless of its low degree of combinability with adjectives), but our findings show

that *extra* plays a marginal role in INTENSIFICATION, as almost all of its occurrences convey a non-evaluative meaning. The observations related to the restricted use of *extra* in intensifying contexts are likewise apparent in French, where it is posited that, analogously to the situation in Italian, the broader proliferation of *extra* is constrained by the existence of other synonymous prefixes, namely *super* and *hyper* (Izert, 2012).

3 Morphological productivity as a function of semantics: a case study

In addition to quantifying the morphological productivity of intensifying prefixes, *InTens* can serve to illustrate the enhanced insights gained by analyzing productivity for a specific semantic function via an annotated dataset. To that end, we find it useful to contrast the prefixes' productivity within the intensifying domain with their broader productivity in generating derivatives of "all senses".

The popularity of studies on morphological productivity, as highlighted by Dal and Namer (2016), comes from the idea that morphology essentially *equals* productivity. Indeed, should morphological theory be concerned solely with processes of word formation that are productive, then identifying which processes are productive and which are not becomes a central concern in morphological research (Baayen and Lieber, 1991). In the quantitative approach, morphological productivity is conceptualized as fluctuating in the frequency of new coinages. In essence, a process is considered more or less productive based on the number of lexemes it generates (Fernández-Domínguez, 2013).

To quantify the productivity of prefixes, a comprehensive study based on type-token ratio (*TTR*), potential productivity ($\mathscr{P}$) (Baayen, 2009), entropy (H) (Shannon, 1948), and population vocabulary size (S) from the finite Zipf-Mandelbrot model for LNRE (Evert and Baroni, 2007; Baroni and Evert, 2014) was conducted. Each measure captures distinct aspects of word formation and usage. The *TTR* measures balance in usage, $\mathscr{P}$ estimates the likelihood of encountering a new type, H can be seen as a measure of unpredictability in the type-frequency distribution, while the population vocabulary size S extrapolates from the observed sample to a sample of arbitrary size, estimating the total number of types that would be observed if the entire population were sampled (Zeldes, 2012). Since "the main upshot of these debates [on morphological productivity] has been the insight that rather than relying on one measure of productivity exclusively, multiple variables should be taken into account and compared" (Hartmann, 2018, p. 84), we believe that using four measures of productivity is a well-grounded decision.

To enable an unbiased comparison of productivity metrics across different prefixes, while avoiding confounding effects introduced by varying token sizes, it is imperative to establish a uniform sample size. Therefore, for both the broader "all senses" dataset, encompassing derivatives of all semantic functions, and the intensified dataset, a fixed number of tokens were randomly chosen for each prefix. In the case of the "all senses" dataset, a sample size of $1,000$ tokens was used, ensuring sufficient overlap for the smallest group, namely the prefix *arci*. For the intensified dataset, the sample size was set at 635 tokens, representing approximately 70% of the dataset for *extra*, the least frequent prefix. We recognize that varying the sample size might influence productivity correlations, but due to practical constraints, testing different sizes was impractical. Larger samples could not be analyzed since the least frequent prefix has only $1,318$ tokens, and much smaller samples would risk introducing too much variability. With all six prefix groups standardized to the same token size, we computed the four productivity measures. To ensure the robustness of our results, this process was iterated 100 times, drawing new random samples (taken `with replacement`) in each iteration.

Figure 1 represents the median values of the four productivity measures for both "all senses" and intensified datasets.

Visual inspection of the plot shows a general trend of increase in the type-token ratio (*TTR*) and potential productivity ($\mathscr{P}$) when prefixes are used to create intensified derivatives. The only exception to this pattern is *extra*, which shows a slight decrease in $\mathscr{P}$. The most significant positive variation in the two measures is apparent with *super*. On the other hand, *stra* and *arci* exhibit minimal changes across all measures. As previously noted, this consistency suggests that the two are predominantly used within the intensifying domain and, therefore, their limitation to intensifying contexts does not notably affect their productivity, as this is their principal area of application. Although variations in *TTR* and $\mathscr{P}$

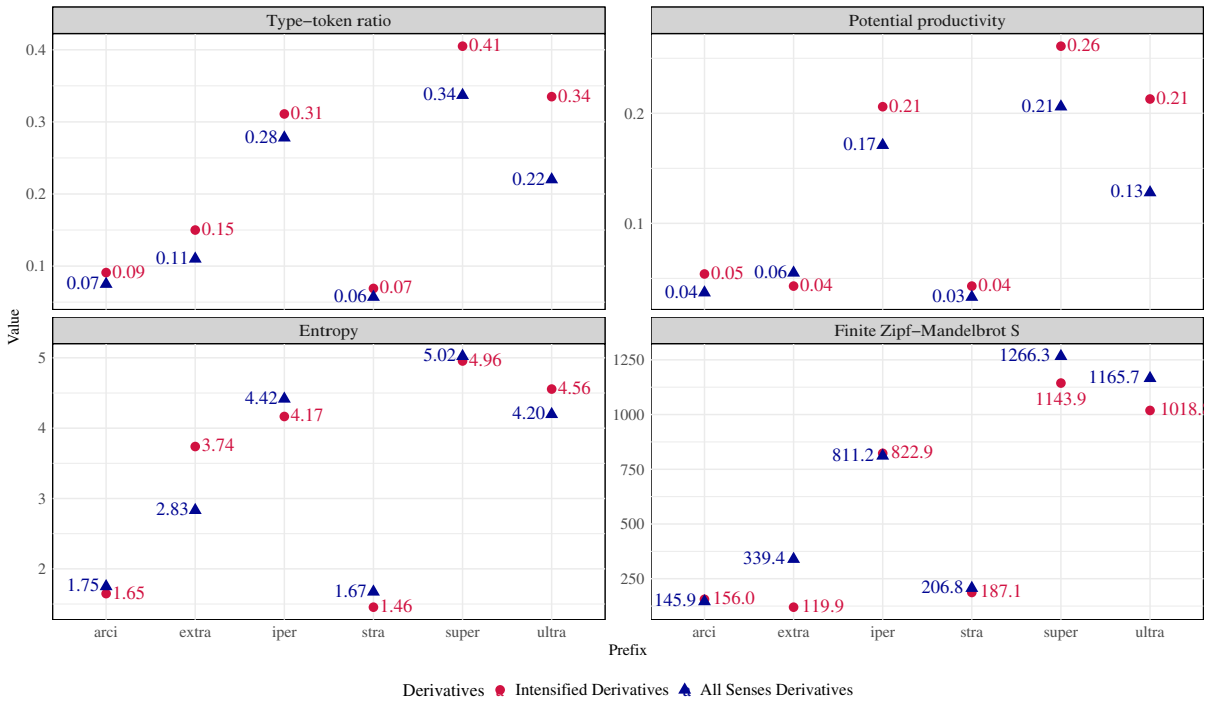

Figure 1: Median productivity values for six prefixes forming derivatives of all semantic functions (represented as triangles) and solely intensified derivatives (represented as dots).

are consistent across all prefixes except for *extra*, larger discrepancies can be seen for entropy (H) and population vocabulary size (S). The observed non-uniformity in H changes indicates alterations in the predictability of the derivative distributions associated with each prefix. For example, the noteworthy increase in H observed for *extra* can be ascribed to the prevalence of frequent, non-intensified derivatives within its general usage. When these highly predictable, frequent forms are excluded to concentrate exclusively on intensified derivatives, the remaining, less frequent items result in a distribution where the occurrence of any particular form becomes less predictable, thereby increasing H. In contrast, the most significant reduction in H is observed for *iper*. This prefix is inclined to produce numerous less frequent, often scientific terms. When these infrequent derivatives are omitted by focusing solely on intensified forms, the ensuing distribution is governed by these more frequent forms, resulting in increased predictability and a reduction in H.

In summary, all measures demonstrate that *super*, *ultra*, and *iper* are the most productive prefixes in the formation of both derivatives of "all senses" and purely intensified derivatives. Notably, these findings align with those of Cartier and Huyghe (2021) concerning French, where *hyper*, *ultra*, and *super* were identified as the most productive high-degree prefixes. Additionally, the same findings can be associated with the perception of *super* as the least intense intensifying prefix in Italian (Calpestrati, 2017). Indeed, should the hypothesis that intensifiers lose their efficacy due to overuse (Mutz, 2015) be accurate, then *super*'s high productivity and low perceived intensity have a strong correlation. On the other hand, *arci* and *stra* exhibit notably low productivity, likely because a single type dominates a large portion of their samples (65% for *arci* with *arcinoto* 'very well-known', and 58% for *stra* with *stragrande* 'vast'). This finding supports the common assertion that low-productivity categories often contain a multitude of high-frequency forms (Plag, 2003).

To assess whether the six prefixes' productivity differs when forming intensified derivatives versus those of all semantic functions, while acknowledging that the intensified dataset constitutes a subset of the comprehensive "all senses" superset, we opted for a permutation test. For each prefix, we compared the observed productivity of its intensified derivatives (calculated from 635 tokens, as previously introduced) against a null distribution generated by repeatedly drawing $1,000$ random samples (also 635 tokens,

`with replacement`) from the "all senses" for that same prefix. For each random sample, we calculated the aforecited productivity measures and then derived an empirical two-tailed p-value, indicating the probability of observing a median productivity as extreme as our actual intensified median purely by chance. The permutation results, completed with Holm-Bonferroni adjusted p-values and Cohen's d effect sizes, revealed statistically significant differences in productivity profiles for certain prefixes.

Prefixes *arci*, *iper*, and *super* showed no statistically significant differences across any measure (all $p_{\text{adj}} = 1$). This indicates that for these prefixes, the observed productivity of their intensified derivatives is not statistically distinct from what would be expected by random chance from their overall usage, implying that intensification does not uniquely impact their productivity in a robust way. In contrast, *extra* and *ultra* exhibited multiple highly significant differences. For *extra*, H ($p_{\text{adj}} = 0$) was significant, with a very large effect size ($d = 11.60$). On the other hand, *ultra*'s intensified derivatives demonstrated highly significant differences in TTR ($p_{\text{adj}} = 0$, $d = 5.16$), $\mathscr{P}$ ($p_{\text{adj}} = 0$, $d = 3.64$), and H ($p_{\text{adj}} = 0$, $d = 6.15$). The large magnitudes of these effect sizes underscore the practical significance of these findings, indicating that the intensified uses of *extra* and *ultra* are significantly more productive in these specific aspects than would be expected from their overall derivative pools. Finally, *stra* showed a large negative effect for H ($d = -2.94$), indicating a lower diversity in its intensified forms. However, this difference is not significant after correction ($p_{\text{adj}} = 0.08$).

Overall, our findings suggest that productivity differences of the prefixes involved in the formation of intensified derivatives versus those of all semantic functions do deviate in a statistically robust way, though not uniformly across all prefixes. Methodologically, these results corroborate the need for extensive semantic annotation in affix rivalry studies, as it allows for a clearer distinction between prefixes' various uses across semantic functions, providing a more precise approach to this complex, gradient phenomenon.

4 Concluding remarks

In this study, we outlined a methodology for building a small-scale dataset of intensified Italian derivatives. This work highlighted the critical role of semantic annotation for polyfunctional derivatives, particularly in the context of affix rivalry. Our subsequent illustrative analysis then demonstrated how the productivity of certain prefixes notably changed when the domain of interest was narrowed from "all senses" to the specific semantic function of INTENSIFICATION.

Beyond the presented (mostly methodological) insights and building upon findings of Section 3, our findings compel a reflection on the intricate relationship between morphological competition and productivity. As noted by Fernández-Domínguez (2017), literature presents divergent views on whether a morphological process gains productivity as a result of previously prevailing in competitive contexts, or if it prevails in competition owing to a prior enhancement in its productivity. For instance, Scherer (2015) sees the competition as a language-internal factor whose changes cause variations in the productivity of processes, while van Marle (1988) describes competition as a gradual process where a decrease in productivity of one process leads to the rise of a competing process that will eventually become so productive that it will replace the original one. While our brief productivity analysis provided a synchronic snapshot of this relationship, a comprehensive understanding of the underlying mechanisms necessitates a fine-grained evaluation of diachronic productivity. Nonetheless, what seems clear from the analysis is that the six prefixes persist in coexistence, with no single process definitively supplanting another. The three most productive prefixes – *super*, *ultra*, and *iper* – exhibit comparably elevated levels of productivity. In contrast, the comparatively lower productivity of *arci* and *stra*, combined with their ability to generate highly lexicalized derivatives, suggests that these prefixes have started developing "niche productivity" (in the sense of Lindsay and Aronoff (2013)) and they might persist precisely because of these established forms, as affixes depend on a certain inventory of frequent types to remain cognitively present. This phenomenon could influence morphological competition, given that lexicalized lexemes often act as blocking agents. Notably, when examining the bases *noto* 'known' and *grande* 'big', two of the most idiosyncratic bases that *arci* and *stra* combine with, it is observed that although the frequency of *arcinoto* 'very well-known' and *stragrande* 'vast' is exceptionally high, the four other examined prefixes seldom combine with these two bases.

Furthermore, it should be noted that variations in productivity may arise from factors other than competition, such as sociolinguistic and pragmatic variables (Körtvélyessy, 2010; Merlini Barbaresi and Dressler, 2020), as well as language fashion, and, on the other hand, the emergence of new processes that enter in competition with an existing one can be attributed to influences beyond productivity, such as language change and language contact (Nagano et al., 2024). We believe that is the case for our most productive prefix — *super*. According to Migliorini (1963), *super* started its diffusion after the appearance of the leader word *super-uomo* 'superman' (after Nietzsche's *Übermensch*) and has spread quickly through contact with mass media. However, existing theories regarding the nexus between competition and productivity are based on traditional morphological competition, which may not be applicable in our context, since within EM rivalry is controlled differently compared to typical derivational morphology (Grandi, 2023). Indeed, it is necessary to examine the pragmatics of EM as a catalyst for competition and to investigate the interplay between competition and productivity within that framework. Space constraints, however, limit this inquiry.

Ultimately, when interpreting these productivity results from the perspective of actual language use, it is crucial to acknowledge their contingency on the corpus employed. The itWaC corpus, developed between 2005 and 2007, is somewhat outdated, and possible shifts in the productivity of the prefixes over the past two decades cannot be ruled out. Diachronic studies typically describe a continuous trend of increasing use of multiple evaluative prefixes throughout the 19th and 20th centuries (Cartier and Huyghe, 2021), with certain prefixes expanding their use diachronically, which may affect their productivity. Such changes may also manifest in micro-diachrony, namely within the temporal range from the creation of itWaC to the present day, given that intensifiers constitute a rather volatile area of language that is particularly susceptible to linguistic innovations (Tagliamonte, 2016).

All things considered, we believe that the creation of this detailed dataset represents a valuable resource for advancing the analysis of prefixal intensifiers in Italian. It provides a foundation for a wide range of qualitative and quantitative investigations, all of which are essential to gain a more nuanced understanding of the complex and multifactorial phenomenon of affix rivalry.

Acknowledgments

I am grateful to Olivier Bonami and three anonymous reviewers for their insightful comments on earlier drafts of this manuscript.

References

Ron Artstein. 2017. Inter-annotator agreement. In Nancy Ide and James Pustejovsky, editors, *Handbook of Linguistic Annotation*, Springer, pages 297–313. https://doi.org/10.1007/978-94-024-0881-2_11.

Harald Baayen. 2009. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling and Merja Kyto, editors, *Corpus Linguistics: An International Handbook*, De Gruyter, pages 899–919. https://doi.org/10.1515/9783110213881.2.899.

Harald Baayen and Rochelle Lieber. 1991. Productivity and English word-formations: A corpus-based study. *Linguistics* 29(5):801–844. https://doi.org/10.1515/ling.1991.29.5.801.

Marco Baroni, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. 2009. The wacky wide web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation* 43:209–226. https://doi.org/10.1007/s10579-009-9081-4.

Marco Baroni and Stefan Evert. 2014. The zipfR package for lexical statistics: A tutorial introduction. https://zipfr.r-forge.r-project.org/materials/zipfr-tutorial.pdf.

Olivier Bonami and Juliette Thuilier. 2019. A statistical approach to rivalry in lexeme formation: French *-iser* and *-ifier*. *Word Structure* 12(1):4–41. https://doi.org/10.3366/word.2018.0130.

Vaclav Brezina. 2018. *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge University Press. https://doi.org/10.1017/9781316410899.

Nicolò Calpestrati. 2017. Intensification strategies in German and Italian written language. In Maria Napoli and Miriam Ravetto, editors, *Exploring Intensification: Synchronic, Diachronic & Cross-Linguistic Perspectives*, John Benjamins, pages 305–326. https://doi.org/10.1075/slcs.189.16cal.

Emmanuel Cartier and Richard Huyghe. 2021. La concurrence affixale en diachronie: Le cas des préfixes de haut degré en français. *Linx* 82. https://doi.org/10.4000/linx.8078.

Georgette Dal and Fiammetta Namer. 2016. Productivity. In Andrew Hippisley and Gregory Stump, editors, *The Cambridge Handbook of Morphology*, Cambridge University Press, pages 70–90. https://doi.org/10.1017/9781139814720.004.

Wolfgang U. Dressler, Lavinia Merlini Barbaresi, Sonja Schwaiger, Jutta Ransmayr, Sabine Sommer-Lolei, and Katharina Korecky-Kröll. 2019. Rivalry and lack of blocking among Italian and German diminutives in adult and child language. In Franz Rainer, Francesco Gardani, Wolfgang U. Dressler, and Hans Christian Luschützky, editors, *Competition in Inflection and Word-Formation*, Springer, pages 123–143. https://doi.org/10.1007/978-3-030-02550-2_5.

Angeliki Efthymiou, Georgia Fragaki, and Angelos Markos. 2015. 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. https://doi.org/10.1556/064.2015.62.4.4.

Stefan Evert and Marco Baroni. 2007. zipfR: Word frequency distributions in R. In *Proceedings of the 45th Annual Meeting of the ACL, Demo and Poster Sessions*. pages 29–32. https://aclanthology.org/P07-2008/.

Jesús Fernández-Domínguez. 2013. Morphological productivity measurement: Exploring qualitative versus quantitative approaches. *English Studies* 94(4):422–447. https://doi.org/10.1080/0013838X.2013.780823.

Jesús Fernández-Domínguez. 2017. Methodological and procedural issues in the quantification of morphological competition. In Juan Santana-Lario and Salvador Valera-Hernández, editors, *Competing Patterns in English Affixation*, Peter Lang, pages 67–117.

Joseph L. Fleiss. 1981. *Statistical Methods for Rates and Proportions.* Second edition. John Wiley & Sons. https://doi.org/10.1002/0471445428.

Bernard Fradin, Georgette Dal, Natalia Grabar, Stephanie Lignon, Fiammetta Namer, Delphine Tribout, and Pierre Zweigenbaum. 2008. Remarques sur l'usage des corpus en morphologie. *Langages* 171(3):34–59. https://doi.org/10.3917/lang.171.0034.

Livio Gaeta and Davide Ricca. 2006. Productivity in Italian word formation: A variable-corpus approach. *Linguistics* 41:57–89. https://doi.org/10.1515/LING.2006.003.

Livio Gaeta and Davide Ricca. 2015. Productivity. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen, and Franz Rainer, editors, *Word-Formation: An International Handbook of the Languages of Europe–Vol. 2*, De Gruyter, pages 842–858. https://doi.org/10.1515/9783110246278-003.

Francesco Gardani, Franz Rainer, and Hans Christian Luschützky. 2019. Competition in morphology: A historical outline. In Franz Rainer, Francesco Gardani, Wolfgang U. Dressler, and Hans Christian Luschützky, editors, *Competition in Inflection and Word-Formation*, Springer, pages 3–36. https://doi.org/10.1007/978-3-030-02550-2_1.

Leo A. Goodman and William H. Kruskal. 1959. Measures of association for cross classifications. II: Further discussion and references. *Journal of the American Statistical Association* 54(285):123–163. https://doi.org/10.1080/01621459.1959.10501503.

Nicola Grandi. 2015. The place of evaluation within morphology. In Nicola Grandi and Lívia Körtvélyessy, editors, *Edinburgh Handbook of Evaluative Morphology*, Edinburgh University Press, pages 74–90. https://doi.org/10.1515/9780748681754-010.

Nicola Grandi. 2023. Evaluative morphology in the Romance languages. In Michele Loporcaro, Francesco Gardani, Patricia Cabredo Hofherr, Jeroen Claes, Andreas Dufter, Martin Maiden, and Franz Rainer, editors, *Oxford Research Encyclopedia of Linguistics*, Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.684.

Matías Guzmán Naranjo and Olivier Bonami. 2023. A distributional assessment of rivalry in word formation. *Word Structure* 16(1):87–114. https://doi.org/10.3366/word.2023.0222.

Kilem Li Gwet. 2008. Computing inter-rater reliability and its variance in the presence of high agreement. *British Journal of Mathematical and Statistical Psychology* 61(1):29–48. https://doi.org/10.1348/000711006X126600.

Stefan Hartmann. 2018. Derivational morphology in flux: A case study of word-formation change in German. *Cognitive Linguistics* 29(1):77–119. https://doi.org/10.1515/cog-2016-0146.

Richard Huyghe and Rossella Varvara. 2023. Affix rivalry: Theoretical and methodological challenges. *Word Structure* 16(1):1–23. https://doi.org/10.3366/word.2023.0218.

Claudio Iacobini. 2004. Prefissazione. In Maria Gorssman and Franz Rainer, editors, *La Formazione delle Parole in Italiano*, De Gruyter, pages 97–163.

Rika Ito and Sali Tagliamonte. 2003. *Well* weird, *right* dodgy, *very* strange, *really* cool: Layering and recycling in English intensifiers. *Language in Society* 32(2):257–279. https://doi.org/10.1017/S0047404503322055.

Małgorzata Izert. 2012. Préfixes *extra-* et *supra-* comme intensificateurs de la propriété en français contemporain. *Kwartalnik Neofilologiczny* 59:437–446. https://journals.pan.pl/publication/102550/edition/88565/kwartalnik-neofilologiczny-2012-no-4.

Dieter Kastovsky. 1986. The problem of productivity in word formation. *Linguistics* 24(3):585–600. https://doi.org/10.1515/ling.1986.24.3.585.

Lívia Körtvélyessy. 2010. *Vplyv Sociolingvistických Faktorov na Produktivitu v Slovotvorbe [On the Influence of Sociolinguistic Factors upon Productivity in Word-Formation]*. SLOVACONTACT.

Lívia Körtvélyessy. 2015. Evaluative morphology and language universals. In Nicola Grandi and Lívia Körtvélyessy, editors, *Edinburgh Handbook of Evaluative Morphology*, Edinburgh University Press, pages 61–73. https://doi.org/10.1515/9780748681754-009.

Mark Lindsay and Mark Aronoff. 2013. Natural selection in self-organizing morphological systems. In Nabil Hathout, Fabio Montermini, and Jesse Tseng, editors, *Morphology in Toulouse: Selected prodeedings of Décembrettes 7*. pages 133–153.

Caterina Mauri, Silvia Ballarè, Eugenio Goria, Massimo Cerruti, and Francesco Suriano. 2019. KIParla corpus: A new resource for spoken Italian. In Raffaella Bernardi, Roberto Navigli, and Giovanni Semeraro, editors, *CLiC-it 2019 – Proceedings of the Sixth Italian Conference on Computational Linguistics*. CEUR.

Lavinia Merlini Barbaresi and Wolfgang U. Dressler. 2020. Pragmatic explanations in morphology. In Vito Pirrelli, Ingo Plag, and Wolfgang U. Dressler, editors, *Word Knowledge and Word Usage: A Cross-Disciplinary Guide to the Mental Lexicon*, De Gruyter, volume 405–452, pages 405–451. https://doi.org/10.1515/9783110440577-011.

Bruno Migliorini. 1963. Fortuna del prefisso *super-*. In Bruno Migliorini, editor, *Saggi sulla Lingua del Novecento*, Sansoni, pages 61–69.

Fabio Montermini. 2008. *Il Lato Sinistro della Morfologia: La Prefissazione in Italiano e nelle Lingue del Mondo*. FrancoAngeli.

Katrin Mutz. 2015. Evaluative morphology in a diachronic perspective. In Nicola Grandi and Lívia Körtvélyessy, editors, *Edinburgh Handbook of Evaluative Morphology*, Edinburgh University Press, pages 142–154. https://doi.org/10.1515/9780748681754-015.

Akiko Nagano, Alexandra Bagasheva, and Vincent Renner. 2024. Towards a competition-based word-formation theory. In Alexandra Bagasheva, Akiko Nagano, and Vincent Renner, editors, *Competition in Word-Formation*, John Benjamins, pages 1–31. https://doi.org/10.1075/la.284.01nag.

Fiammetta Namer, Nabil Hathout, Dany Amiot, Lucie Barque, Olivier Bonami, Gilles Boyé, Basilio Calderone, Julie Cattini, Georgette Dal, et al. 2023. Démonette-2: A derivational database for French with broad lexical coverage and fine-grained morphological descriptions. *Lexique* 33:6–40. https://doi.org/ff10.54563/lexique.1242.

Maria Napoli. 2012. Uno stra-prefisso: L'evoluzione di *stra-* nella storia dell'italiano. *Rivista Italiana di Linguistica e Dialettologia* 14:89–112.

Ingo Plag. 1999. *Morphological productivity: Structural constraints in English derivation*. De Gruyter. https://doi.org/10.1515/9783110802863.

Ingo Plag. 2003. *Word-Formation in English*. Cambridge University Press. https://doi.org/10.1017/9781316771402.

Carmen Scherer. 2015. Change in productivity. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen, and Franz Rainer, editors, *Word-Formation: An International Handbook of the Languages of Europe–Vol. 3*, De Gruyter, pages 1781–1793. https://doi.org/10.1515/9783110375732-014.

Claude E. Shannon. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27(3):379–423.

Paulo Sergio Panse Silveira and Jose Oliveira Siqueira. 2023. Better to be in agreement than in bad company: A critical analysis of many kappa-like tests. *Behavior Research Methods* 55(7):3326–3347. https://doi.org/10.3758/s13428-022-01950-0.

Sali Tagliamonte. 2016. *Teen Talk: The Language of Adolescents*. Cambridge University Press. https://doi.org/10.1017/CBO9781139583800.

Werner Vach and Oke Gerke. 2023. Gwet's AC1 is not a substitute for Cohen's kappa: A comparison of basic properties. *MethodsX* 10:102212. https://doi.org/10.1016/j.mex.2023.102212.

Jaap van Marle. 1988. On the role of semantics in productivity change. In Geerd Booij and Jaap van Marle, editors, *Yearbook of Morphology 1988*, Foris, pages 139–154. https://doi.org/10.1007/978-94-017-3710-4_7.

Tatu Ylonen. 2022. Wiktextract: Wiktionary as machine-readable structured data. In Nicoletta Calzolari et al., editors, *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC)*. pages 1317–1325. https://aclanthology.org/2022.lrec-1.140.

Eros Zanchetta and Marco Baroni. 2005. Morph-it! A free corpus-based morphological resource for the Italian language. In *Proceedings from the Corpus Linguistics Conference Series 2005*. University of Birmingham. https://hdl.handle.net/11585/15321.

Amir Zeldes. 2012. *Productivity in Argument Selection: From Morphology to Syntax*. De Gruyter. https://doi.org/10.1515/9783110303919.

Nicola Zingarelli, Mario Cannella, Beata Lazzarini, and Andrea Zaninello. 2023. *Lo Zingarelli 2024: Vocabolario della Lingua Italiana*. Zanichelli.