Rossella Varvara

Constraints on nominalizations: Investigating the productivity domain of Italian *-mento* and *-zione*

Abstract: The paper investigates the different productivity domains (Rainer 2005) of two Italian event denoting suffixes, *-mento* and *-zione*. These suffixes share the same eventive semantics, they are both productive and thus can be seen as rivals in the formation of event nominalizations. The aim is to obtain a better understanding of the constraints that play a role in the selection of one affix over the other. By means of a logistic regression model the contribution of different features of the base verb is investigated. The analysis is conducted on a dataset of 678 nominalizations extracted from a section of Midia, a diachronic balanced corpus explicitly built for morphological research (Gaeta 2017). Results show that the frequency, the inflectional class and the number of characters of the base verb as well as the presence of the prefix *a-* significantly contribute to the definition of the different domains, only partially confirming previous findings.

Keywords: constraints on productivity, competition, derivational morphology, action nouns, deverbal nominalizations, corpus linguistics, quantitative analysis

1. Introduction

Competition among affixes (also known as *rivalry*) is a common phenomenon in morphology (see a.o. Rainer et al. 2019), both in inflection and word-formation. It can be described as the availability of multiple patterns to express a certain concept. Research on this topic usually focuses on understanding which form is preferred by a speaker or by a speech community and what are the reasons underlying this choice.

The competing patterns may differ in their degree of productivity[1]: one could be more available to form neologisms in the present-day language than the other. Or they could be productive in different domains, i.e. different subsets of words to which the pattern applies (Rainer 2005). The pattern's domain can be defined through the features (called *constraints* or *restrictions*) that a potential base should possess. An example comes from English nominalizations in *-ation*: contrary to the *-al, -ance, -ment* or *-ure* nominalizations, they apply to suffixed verbs in *-ize* and *-ify* (e.g. *adultification, aristocratization,* Plag 2003: 63, Bauer et al. 2013: 196–202).

---

1   For an overview of the notion and measures of *productivity* see Bauer (2001, 2005) and Baayen (2009).

Constraints can be of different nature, as they can concern phonological, morphological or semantic aspects of the base. An example of a morphological condition has just been described above for the English suffix *-ation*, whereas we can cite as a case of phonological constraint the preference of the English suffix *-eer* for bases ending in [t] (e.g. *musketeer*, *profiteer*, *racketeer*, Adams 1973: 175-178). A syntactic restriction is at play in the preference of the suffix *-able* for transitive base verbs: *visitable* vs **goable*, *observable* vs **lookable* (Rainer 2005: 348). At the semantic level, the Spanish relational suffix *-uno* is mostly attached to base nouns referring to animals (e.g. *vaca* 'cow', *vacuno* 'relating to cow'). Among others, Rainer (2005) and Gaeta (2015) offer overviews of the different types of constraints a pattern may present, with further examples from multiple languages.

The present paper focuses on the differences in the productivity domain of two Italian competing patterns, i.e. nominalizations in *-mento* and *-zione*. They belong to a specific class of nominalizations, here called *event-denoting deverbal nominalizations* (henceforth EDN). The term nominalization indicates both the process and the result of "turning something into a noun" (Comrie and Thompson 2007: 334), but in this context we restrict our analysis to cases in which the base of the process is a verb and the resulting nominalization refers to an event (in the broadest sense[2]).

In many languages, more than one affix is available to form an EDN. In English, for example, the suffixes *-al* (*arrival, approval*), *-ance* (*resistance, attendance*), *-ing* (*reading, learning*), *-ation* (*regulation, consultation*), and *-ment* (*recruitment, development*) can all be used to form event nominalizations. These can be seen as constituting a single paradigmatic cell of semantic derivation (Booij and Lieber 2004). In Italian, the language under investigation in this work, multiple suffixes are available to exploit this function as well: *-zione* (*venerazione* 'veneration'), *-mento* (*annegamento* 'drowning'), *-tura* (*spuntatura* 'trim'), *-aggio* (*smontaggio* 'dismanteling'), *-ata* (*sbirciata* 'peek'), *-nza* (*permanenza* 'permanence, stay'). Moreover, event nouns may be formed also by means of conversion (or *zero derivation*): *aumento* ('increase'), *viaggio* ('trip'). Among all these patterns, the *-mento* and *-zione* have been selected for this first investigation since they are the most productive. The study aims at understanding if some morphological properties of the base verbs are relevant in the selection of one affix instead of the other in the formation of EDNs.

---

2 With the term "event" I refer to every kind of eventuality (Bach 1986), including states. Thus, event nominalizations may denote activities, achievements, accomplishments and states (following the terminology proposed in Vendler 1957). The same class of derived nominals has been frequently called *action nouns* (or *nomina actionis,* Comrie 1976; Comrie and Thompson 2007; Koptjevskaja-Tamm 1993, 2006).

The analysis is conducted by considering all the formations attested in a corpus from a specific period of time (from 1841 to 1947), thus investigating the constraints on *realized productivity* (Baayen 2009), rather than on *expanding* or *potential productivity*. The realized productivity (also known as *extent of use*) measures the number of complex words a morphological process produced in the past. Conversely, expanding and potential productivity are seen as measures of the expansion of the class in the near future, i.e. how much the morphological processes are expected to be used to form neologisms. In this study, the analysis is conducted on "past" formations, but future work can investigate the productivity constraints considering only neologisms in the dataset.

The paper is structured as follows. In the next section (§2), I introduce previous findings on the differential constraints of *-mento* and *-zione*. In section 3, I present the methodology applied, i.e. regression modelling based on corpus data, and the data sampling. In section 4, I list the variables considered as predictors and provide their descriptive statistics. In section 5, the regression modelling and its results are presented. Section 6 offers a discussion of the main findings. Section 7 draws conclusions and directions of future research.

2. Previous works on Italian *-mento* and *-zione*

Numerous works have focused on assessing the productivity degree of the suffixes *-mento* and *-zione* (Thornton 1988; Iacobini & Thornton 1992; Gaeta & Ricca 2002; Fiorentino 2008; Štichauer 2009; Varvara 2019), and they frequently considered also other EDNs. Similarly, the problem of the stem form to which they attach has often been discussed (Scalise 1983; Thornton 1990–1991, 2015; Gaeta 2004). The issue of the competition between these two Italian suffixes has been addressed by Scalise (1983: 207–208), Melloni (2007: 70–71), and in more depth by Gaeta (2002 2004: 327 and ff). They propose numerous constraints on the productivity of *-zione* and *-mento* suffixes. For an easier description, we can divide them in phonological, morphological and semantic constraints.

2.1. Phonological constraints

Gaeta (2004) observes that *-zione* is used to derive action nouns for some of the few Italian monosyllabic verbs (ex. *dizione* 'diction', from *dire* 'to say', *stazione* 'station', from *stare* 'to stay', *dazione* 'dation' from *dare* 'to give'), whereas *-mento* attaches to bases that are at least bisyllabic. Second, he notices a euphonic restriction for *-zione*, which usually does not follow bases that end in /ts/ plus a vowel (e.g. **deprezzazione* from the base verb *deprezzare* 'depreciate').

2.2. Morphological constraints

Previous works do not recognize a specific association with the inflectional class of the base verb; however, parasynthetic verbs from the third conjugation seem to prefer -*mento* (*impigrimento* 'the act of becoming lazy'). Moreover, previous analyses note a slight preference of simple base verbs for -*mento* (e.g. *biascicamento* 'munching'), whereas base verbs formed by conversion prefer -*zione* (e.g. *datazione* 'dating').

Various associations with base verbs that present some specific affixes are listed (Gaeta 2004: 327–331). First, they highlight an association of the suffix -*mento* with parasynthetic verbs formed with the prefixes *ad-* and *in-* (*ammanettamento* 'handcuffing', *inacidimento* 'souring'). Second, prefixed verbs with *s-* are more correlated with -*mento*. More specifically, this prediction is linked to the two meanings that the prefix *s-* can bring in Italian: a negative value (e.g. *fiorire* 'to bloom' vs *sfiorire* 'to wither') and an intensifier one (e.g. *gridare* 'to shout' vs *sgridare* 'to scold'). Nouns in -*mento* are formed from prefixed bases with either of these two meanings, whereas -*zione* attaches mainly to bases in which the prefix *s-* has a negative value. Thus, the verbs with the intensifier *s-* are associated with -*mento* derivatives. Third, the suffix -*zione* is more frequently associated with prefixed verbs with *e(s)-* (e.g. *eruzione* 'eruption') or *de-* (e.g. *decomposizione* 'decomposition'), no matter if the bases are parasynthetic verbs or simple prefixed ones. Fourth, suffixed verbs with -*ific-* and -*izz-* seem to be more frequently associated with derivatives in -*zione (*e.g. *laicizzazione* 'secularization', from *laicizzare* 'to secularize', and *nazificazione* 'nazification', from *nazificare*), even if some derivatives in -*mento* are attested (*volgarizzamento* 'translation into vernacular'). Note, however, that this constraint seems in contradiction with the euphonic restriction we have listed in 2.1., i.e. -*zione* derivatives tend to avoid base verbs ending in /ts/ plus a vowel. Fifth, suffixed verbs in -*eggi-, -acchi-/-ucchi-, -(er)ell-, -ett-, -icch-* prefer -*mento* to form nominalizations (*fronteggiamento* 'confrontation', *saltellamento* 'hopping', *scoppiettamento* 'crackling', *mordicchiamento* 'nibbling'). Lastly, bases with the suffix -*iv-* select the -*zione* suffix (*attivazione* 'activation').

2.3. Semantic constraints

Gaeta (2002: 215 and ff) also identifies a difference in the semantics of the resulting nominalizations. He draws two conclusions: 1- -*mento* derivatives show the simple derivational meaning of 'the act of V' (where V is the base verb) more frequently compared to those in -*zione* (which have thus a higher degree of polysemy); 2- among all the other possible readings, -*zione* shows a high number of derivatives with an additional resultative meaning of 'what has been V-ed'.

3. Methodology

Even if the observations made by Gaeta (2004) were based on quantitative data extracted from the DISC online dictionary and from one year of the *La Stampa* newspaper, the strength of these associations was not assessed, since no statistical test was conducted. As noted by Bonami and Thuilier (2019: 6) "descriptive statistics does not allow one to determine whether the tendencies observed in a sample are robust enough that one can exclude their being due to chance: neither do they allow one to conclude on the relative role of highly correlated properties of the base, such as phonological and morphological characterizations of their shape". For these reasons, raw counts of occurrences in a corpus are not enough to assess the correlation among phenomena; moreover, since there are multiple possible factors at play, a multivariate statistical model is more suitable for this kind of investigation than single monofactorial tests.

Statistical approaches are nowadays widespread in linguistics, and also specifically in the study of the rivalry between affixes. Arndt-Lappe (2014) applies an analogical model to investigate the rivalry between the English suffix *-ity* and *-ness*. Her findings show that the model can predict the preference patterns by the phonological characteristics of the two base-final syllables and by the syntactic category of the base. Varvara (2017) focuses on the competition between Italian nominal infinitives and the whole class of event-denoting nominalization suffixes, by applying a regression analysis to evaluate possible constraints. Bonami and Thuiller (2019) focus on the French suffixes *-iser* and *-ifier*, and they highlight how multiple factors may play a role at the same time.

Similarly to the latter two works, the present study applies logistic regression modelling, inspecting the domain of application of the Italian suffixes *-mento* and *-zione*. In binary logistic regression, the model estimates the probability of a predicted event whose outcome is binary (0 or 1). In our case, the predicted event is the suffix used to form a nominalization from a base verb. This is our dependent variable (also called response) and it has the binary outcome *-mento* vs *-zione*. Given the data observed, the model will assess the role of different predictors (or independent variables, i.e. the constraints investigated) in the selection of the suffix used.

The analysis is conducted on data extracted from the MIDIA corpus[3], a diachronic corpus of 7,5 million tokens. Even if the corpus size is quite small, it has the advantage of being balanced through genres. For the present work, only the period[4] from 1841 to 1947 has been considered as a

---

3   www.corpusmidia.unito.it
4   The subcorpus selected contains 1.667.928 tokens.

first step, but future work will extend the analysis to the whole corpus and compare the results diachronically. From the texts available for this time span, all the occurrences of words in *-mento* and *-zione* have been automatically extracted. As a second step, a manual check was done to remove all the typos and, following a procedure similar to that adopted by Gaeta and Ricca (2002: 233–237), to also remove:

1- simple, not morphologically complex nouns that accidentally end in *-mento* or *-zione* (e.g. *elemento* 'element', *cemento* 'cement');

2- opaque derivations whose semantic relation with the bases is no more transparent (e.g. *stazione* 'station', derived from *stare* 'to stay');

3- denominal nouns (e.g. *tunnellamento*), since the aim of the study is to assess the relevant properties of the base verbs in the suffix selection.

The resulting dataset consisted of 678 items, 249 nouns in *-mento* and 429 nouns in *-zione*. Each lemma was thus annotated with a set of 7 features, i.e. possible constraints on pattern productivity. These variables are listed in the next section.

4. Predictors and descriptive statistics

In this section I describe each variable and present some descriptive statistics. The set of variables considered as predictors are:

- Frequency of the base (continuous variable);
- Frequency of the derived term (continuous variable);
- Ratio of the derivative frequency to the base frequency (continuous variable);
- Length in characters of the base verb (continuous variable with values from 6 to 15);
- Inflectional class of the base verb (categorical, with three levels: *-are*, *-ere*, *-ire)*;
- Number of total derivational processes of the derivative (continuous variable with values from 1 to 4);
- Other affixes present on the base verb (categorical, with 14 levels).

4.1. Control variables

The first variable considered is the frequency of the nominalization in the corpus MIDIA. The frequency distributions of the two categories are slightly different, with *-zione* derivatives in general more frequent than *-mento* ones. Tab. 1 reports their token frequency in the corpus (providing minimum and maximum values, median and mean).

Tab. 1: Frequency distributions for -mento and -zione nominals.

	Min.	Median	Mean	Max
-mento	1	4	17.58	409
-zione	1	7	19.64	422

In addition, I take into consideration the frequency of the corresponding base verb. The base verbs for the two groups show a similar frequency distribution, as reported in Tab. 2. In 27 cases, both a *-mento* and a *-zione* derivative was attested for the same verb.

Tab. 2: Frequency distributions of corresponding base verbs.

	Min.	Median	Mean	Max
-mento	1	74	76.2	152
-zione	1	71	70.41	151

Furthermore, the ratio of the EDN frequency to the base frequency (known as *relative frequency*) is considered. Previous work has indeed highlighted that relative frequency is more correlated to morphological processing, productivity and semantic transparency, compared to absolute frequency (see Hay 2001; Baayen 2009). Higher relative frequency seems to be related to faster processing, higher semantic transparency, and higher productivity. Tab. 3 reports the values for the two EDN categories: nominalizations in *-zione* have higher relative frequency than those in *-mento*; on average, the frequency of *-zione* EDNs is equal to the frequency of the corresponding base, whereas the *-mento* EDNs have on average a lower frequency than that of the base.

Tab. 3: Frequency distribution of ratio of EDN frequency to base verb frequency.

	Min.	Median	Mean	Max
-mento	0.0066	0.0758	0.5245	26.2500
-zione	0.0068	0.1467	1.0870	52.3333

4.2. Length in characters of the base verb

Previous accounts (Gaeta 2004; Melloni 2007) found a correlation between monosyllabic bases and *-zione* derivatives. For this reason, the length of the base verb (in the form of the infinitive[5]) is considered as a predictor. In our sample, verbs vary from a minimum of 6 characters to a maximum of 15 (median=9, mean[*-mento*]= 8.976, mean[*-zione*]= 9.177). A more detailed overview is given in Table 4.

---

5   The length is computed in terms of number of characters. However, further analysis may consider the number of phonemes.

Tab. 4: Number of characters of corresponding base verbs.

	Number of characters of the base verb									
	6	7	8	9	10	11	12	13	14	15
-mento	11	30	48	74	49	28	8	0	0	1
-zione	13	45	87	114	86	57	18	7	2	0

4.3. Inflectional class of the base verb

Italian verbs are classified in three main conjugations, depending on the infinitive ending: first conjugation in *-are* (e.g. *mangiare* 'to eat'), second conjugation in *-ere* (*bere* 'to drink'), third conjugation in *-ire* (*sentire* 'to hear'). The distribution of these three classes is reported in Table 5, together with expected values[6]. Previous works (§2) have noticed a relation between the *-mento/-zione* rivalry and the conjugation of the base verb. Specifically, it has been argued that parasynthetic base verbs from the third conjugation more frequently derive EDNs in *-mento*. If we compare observed and expected values, we note that *-mento* derivatives from the third conjugation are higher than expected. This may be due specifically to parasynthetic verbs or to the whole third conjugation. Moreover, the suffix *-mento* associates more frequently than expected with verbs from the second conjugation too, whereas *-zione* associates more with the first conjugation. We will test the significance of these correlations in section 5.

Tab. 5: Inflectional class of base verbs.

	Observed values		Expected values	
	-mento	**-zione**	**-mento**	**-zione**
-are	144	350	181.42	312.58
-ere	46	33	29.01	49.99
-ire	59	46	38.56	66.44

4.4. Morphological complexity

The presence of other morphological processes was considered in two ways: first, by computing the total number of morphological processes; second, by considering the specific affix present. This information was taken, whenever possible, from Derivatario[7] (Talamo, Celata, and Bertinetto 2016), a freely available digital lexicon of morphologically complex Italian words. When the attested lemma was not available in this resource, the annotation was carried out by hand.

---

6   Expected values are computed by means of a chi-squared test.

7   http://derivatario.sns.it

4.4.1. Total number of morphological processes

The total number of morphological processes attested in our sample ranges from 1 to 4 processes. Tab. 6 reports the distribution of EDNs (type frequency) for each number of processes observed. For example, there are 58 nominalizations in -*mento* that show two morphological processes; -*zione* EDNs are instead 102 for this level. One morphological process indicates that the derivative shows only the nominalization process, i.e. it does not contain any other derivational affix besides -*mento* or -*zione*. This is the case of *fondamento* ('foundation'), which is directly derived from the base verb *fondare* ('to found'). The word *armonizzazione* ('harmonization') has undergone two derivational processes: first the formation of the denominal verb *armonizzare* ('to harmonize'), derived from the noun *armonia* ('harmony') by means of the suffix -*izz-*; then, the denominal verb transformation into a deverbal noun by means of the nominalizing suffix -*zione*. An example of a derivative with three derivational processes is *immatricolazione* ('enrolment'): starting from the noun *matricola* ('freshman'), the verb is formed by parasynthesis[8] (e.g. conversion combined with the inchoative prefix *in-*); then, the noun is derived with the suffix -*zione*. For our sample, the maximum number of derivational processes is 4. The word *ristabilimento* ('reinstatement') is an example where the verbal base *stare* ('to stay') is turned into an adjective by means of -*bile*, which is then converted into a verb (*stabilire* 'to establish'), modified by the iterative prefix *re-*, and lastly turned into an EDN by the suffix -*mento*. Table 6 reports additionally the expected values of EDNs in each category (computed by means of a chi-squared test). As can be seen, -*zione* derivatives are more frequent than expected when no other derivational process is present, whereas -*mento* nominalizations occur more than expected when 3 or 4 derivational processes are present. The significance of this difference will be tested in the regression model.

Tab. 6: Number of morphological processes in EDNs.

	Observed values		Expected values	
	-mento	-zione	-mento	-zione
1	127	273	146.90	253.10
2	58	102	58.76	101.24
3	56	54	40.40	69.60
4	8	0	2.94	5.06

---

8   Parasynthesis indeed counts as two derivational processes in the total count.

4.4.2. *Presence of other affixes*

The last variable taken into consideration is the nature of other affixes (whenever present). Tab. 7 reports the different affixes attested in the sample with the type frequency of the two nominalization patterns. Following Talamo, Celata, and Bertinetto (2016), some affixes are split in two groups based on their semantics: e.g. 1de- indicates the prefix *de-* when occurring with a reversative reading (like in *detassare* 'to untax'); 2de- the same prefix with a causative meaning (*depurare* 'to purify'). With the label *a-* I refer to every prefix formed by the sequence *a* plus a consonant (e.g. *abbassare* 'to lower' *avvicinare*, 'to place near'). In the rest of the paper I will call this whole class *prefix a-*. Some affixes are attested only with one nominalization, but the values in some cases are really low and it would not be possible to generalize on these few occurrences. As it will be explained in the next section, in the regression modelling the affixes that have zero formations with one EDN will be aggregated in one level called *other affixes*. Sampling zeros would cause indeed infinite estimates.

Tab. 7: Presence of other affixes.

Affix	Observed values		Affix	Observed values	
	-mento	-zione		-mento	-zione
1 de-	2	4	-eggiare	3	0
1 in-	0	1	-ificare	0	22
1 s-	3	1	-izzare	0	11
2 de-	3	8	-nte	0	1
2 in-	15	14	No other affix	147	310
2 s-	7	1	pre-	0	3
a-	39	12	pro-	3	2
-bile	2	0	ri-	16	8
co-	1	5	trans-	1	3
con-	2	9	-zione	1	0
dis-	3	7			

5. Multifactorial statistical analysis

In summary, we have 7 independent variables in our statistical analysis. Two of them are categorical variables, whereas the other 5 are continuous variables. The response of our model is a binary variable 0/1, where 0 corresponds to the suffix *-mento* and 1 to the suffix *-zione*. All the frequency variables were log-transformed and scaled on their mean. The analyses are performed with the software R (R Core Team 2015), and by means of the *glm* function (with family type equal to binomial).

In order to determine the model that best fits our observations and to keep only significant variables, I proceed with backward selection of the variables, using the *step* function. Given the most complex model (with all the variables together), this function compares it to all the possible alternative models removing one variable at a time and evaluates the best-fitting model based on likelihood ratio tests and AIC[9] values (i.e., the lower, the better). It tries to take into account as much variance as accounted for by the complex model, while removing predictors that do not contribute to the regression equation.

In Tab. 8, the result of this procedure is summarized: in the first row the starting complex model is defined, whereas in the second row the final model is reported.

Tab. 8: Model selection.

	Model	**Deviance**	**AIC**
full model	Freq. EDN + Freq. Base + Rel. freq. + Conjugation + Affixes + Num. of deriv. processes + Base length	732.46	774.46
best model	Freq. EDN + Conjugation + Affixes + Num. of deriv. processes + Base length	733.50	771.5

As shown, the frequency of the base and the relative frequency were not improving the model significantly and were removed. Thus, the predictors of the final model are thus the frequency of the derived word, the inflection class of the base verb, the other affixes present on the base, the total number of derivational processes and the length in characters of the base verb. The probability of deriving a *-zione* derivates is computed based on these variables.

Note, however, that from the original dataset I removed 8 levels of the factor *Other affixes* (specifically 1in-, -bile, -eggiare, -ificare, -izzare, -nte, pre-, -zione), because they presented *sampling zeros* (Agresti 2003: 138). As can be noted in Tab. 7, in our sample these affixes occur only with one of the two nominalizations, and thus show a value of zero in the other cell. Sampling zeros may cause computation problems in the model (with infinite estimates occurring for that level). For this reason, I aggregated these 8 levels into one called *others*. We cannot know if these zeros indicate a true correlation or are either due to sparse data. We can only note that, given our sample and the values reported in Tab. 7, base verbs formed with the suffixes *-eggiare*, *-zione* and *-bile* show only nominalizations in *-mento*,

---

9   Akaike Information Criterion. It is a badness of fit indice and for this reason we pick the model with the lowest index.

whereas base verbs formed with the affixes *1in-, pre-, -ificare, -izzare* and *-nte* present only nominalizations in *-zione*. Further evidence from a larger corpus is needed to confirm these tendencies.

The summary of the final model is reported in Tab. 9. The first column reports the predictors of the model, with a row for each level of the categorical factors. In the second column, the sign of the estimated coefficient[10] indicates the direction of the effect: a positive estimate indicates an association between the factor and the nominalization in *-zione;* a negative one an association with *-mento.* Column 5 shows the significance of a predictor (i.e., its p-value), which indicates how much it contributes to the distinction. The significance level of the p-value is set to < 0.05.

Tab. 9: Summary of the final model.

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	1.372	1.211	1.134	0.256889	
scale(log(Fq_EDN))	0.427	0.098	4.364	1.28e-05	***
conjugation: ere	-1.695	0.291	-5.818	5.97e-09	***
conjugation: ire	-1.421	0.252	-5.631	1.79e-08	***
affix: others	0.272	0.982	0.277	0.781851	
affix: 1S	-1.691	1.546	-1.094	0.274038	
affix: 2DE	0.630	1.133	0.556	0.578213	
affix: 2IN	-0.506	0.970	-0.522	0.601763	
affix: 2S	-2.042	1.392	-1.467	0.142301	
affix: A	-1.908	0.948	-2.013	0.044111	*
affix: CO	0.686	1.446	0.475	0.635039	
affix: CON	0.932	1.184	0.787	0.431039	
affix: DIS	-0.125	1.143	-0.110	0.912667	
affix: No affix	-0.805	0.947	-0.850	0.395277	
affix: PRO	-0.753	1.300	-0.579	0.562748	
affix: RI	-1.215	1.002	-1.213	0.225253	
affix: TRANS	0.315	1.502	0.210	0.833783	
number of derivational processes	-0.728	0.219	-3.327	0.000877	***
length in characters of the base verb	0.172	0.070	2.441	0.014645	*

10 Estimates are expressed in log odds, an alternate way to express probability, whose values range from -∞ to + ∞.

The frequency of the derived term, the inflectional class of the base verb and the number of its derivational processes are all significant predictors (p < 0.001) for the distinction between -*zione* and -*mento* nominalizations.

Specifically, a one-unit increase in frequency is associated with an increase in the log odds of the derivative being a -*zione* nominalization in the amount of 0.427. This finding confirms the trend we already noted in Tab. 1 (p.84) as significantly supported by the statistical data analysis. Nominalizations in -*zione* are more frequently used than those in -*mento*, probably because of their higher productivity.

5.1. Inflectional class of the base verb

With regard to the effect of the inflectional class, we observe that base verbs ending in -*ere* and -*ire* are both associated with the decrease in the log odds for -*zione* nominalizations. In other words, verbs from these conjugation classes tend to derive EDNs with the suffix -*mento*, rather than -*zione*; by contrast, verbs from the first conjugation are more associated with -*zione* nominalizations. Fig. 1 represents this effect[11]. This finding may be interpreted as a correlate of what has been previously observed about parasynthetic verbs in -*ire*, which form nominalizations in -*mento* (§4.3, p.85).
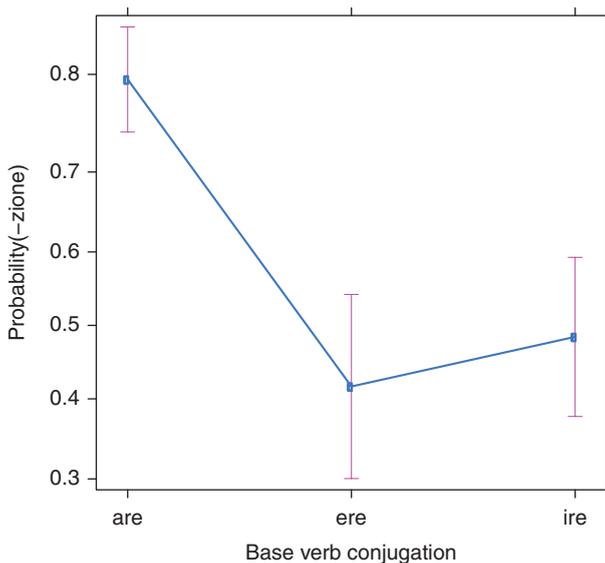

Fig. 1: Inflectional class effect plot.

---

11 Effect plots have been drawn using the R package *effect* (Fox et al. 2019).

Indeed, a quick analysis of verbs in *-ire* forming *-mento* derivatives reveals that 16 of them (out of 59) are parasynthetic verbs, whereas only one parasynthetic verb in *-ire* forms a *-zione* nominalization. However, the role of parasynthetic verbs may be marginal, considering a further association between *-mento* and the Italian second inflectional class of verbs (ending in *-ere*). There is probably something more linked to the base verb conjugation that drives the choice of the nominalizing suffix.

5.2. Total number of morphological processes and length of the base

The total number of morphological processes on the EDN is also significant: a higher number is linked to a decrease of log odds for *-zione* nominalizations. A higher number of morphological processes is thus associated with *-mento* derivatives, fewer processes are associated with *-zione* (Fig. 2). This finding contradicts previous works that stated a slight preference of simple base verbs for *-mento* EDNs.
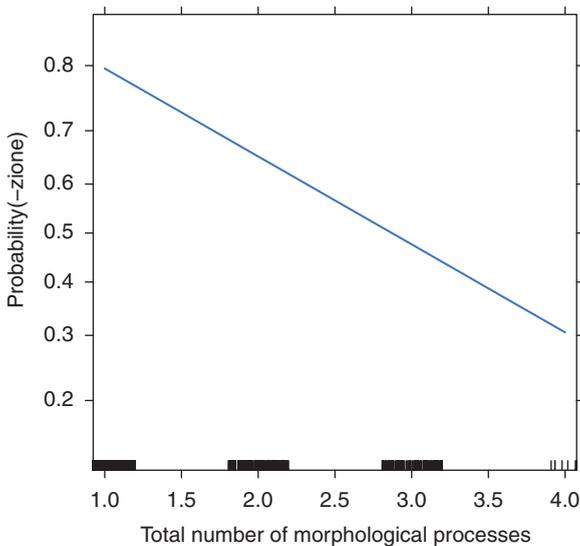

Fig. 2: Number of morphological processes effect.

The number of characters of the base verb is significant as well ($p < 0.05$), but its effect goes in the opposite direction with respect to the number of morphological processes: longer bases are associated with *-zione* EDN (Fig. 3). It is interesting to note that longer bases are thus not correlated with a higher number of derivational processes: *-zione* EDNs are formed from longer bases, but *-mento* EDNs present a higher number of morphological processes. This phenomenon may be linked to the treatment of parasynthetic verbs:
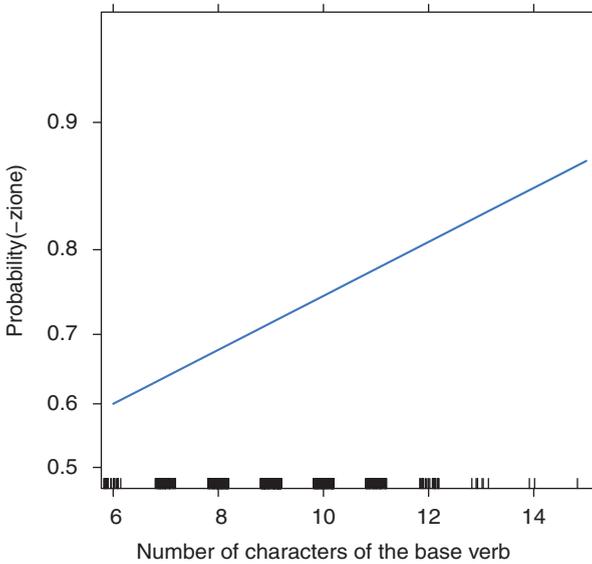
Fig. 3: Effect of the number of characters of the base.

parasynthesis, indeed, has been counted as two morphological processes and EDNs annotated as showing 4 morphological processing (mainly -*mento* EDNs) are indeed all derived from parasynthetic verbs. Nominalizations in -*mento* seem once more associated with this category of verbs.

Still, when considering the number of characters of the base, it should be noted that we do not have monosyllabic verbs (e.g. *dire* 'to say') in our sample and thus we cannot assess the hypothesis (expressed in previous works, see §2.1.) that -*zione* is preferred for these verbs. However, we note that our analysis reveals an opposite tendency, i.e. -*zione* EDNs are associated with longer bases.

5.3. Presence of other affixes

Lastly, the model in Tab. 9 shows that only one level of the variable OTHER AFFIX is significant ($p < 0.05$), i.e. the presence of the prefix *a-*. It reduces the log odds of having a -*zione* derivative, and thus increases those of -*mento* EDNs. Since no other affixes contributed to the model, I decided to simplify the model by considering only the presence (or absence) of the prefix *a-* as factor (instead of the whole list of affixes). The new factor (which I will call *Prefix A*) has three levels: presence (if the base verb shows this affix), no affix

(if the base verb is a simple non-derived verb), other affix (if the base verb has additional affixations besides the prefix *a*-). I then repeated the analysis with a stepwise selection of the significant variables (Tab. 10).

Tab. 10: Model selection considering only the prefix a- *as additional affix.*

	Model	Deviance	AIC
full model	Freq. EDN + Freq. Base + Rel. freq. + Conjugation + Prefix A + Num. of deriv. processes + Base length	750.15	770.15
best model	Freq. EDN + Conjugation + Prefix A + Num. of deriv. processes + Base length	751.29	767.29

The final model contains the same variable as the model from the previous setting (Tab. 8 and Tab. 9). The frequency of the base and the relative frequency were removed in the final model because they were not significantly improving the fit. The summary of this model is reported in Tab. 11.

Tab. 11: Summary of the final model considering only the prefix a- *as additional affix.*

	Estimate	Std. Error	z value	Pr(>lzl)	
(Intercept)	-0.780	0.871	-0.896	0.37049	
Freq. EDN	0.427	0.096	4.437	9.14e-06	***
conjugation -*ere*	-1.824	0.283	-6.448	1.13e-10	***
conjugation -*ire*	-1.504	0.247	-6.083	1.18e-09	***
Prefix A: different affix	1.637	0.395	4.150	3.32e-05	***
Prefix A: No affix	1.119	0.481	2.324	0.02012	*
number of deriv. processes	-0.753	0.206	-3.662	0.00025	***
Base length	0.206	0.067	3.055	0.00225	**

All the effects go in the same direction as in the previous setting; moreover, their significance increased, and the AIC lowered. The effect of the prefix *a*- is confirmed, as the plot in Fig. 4 shows.
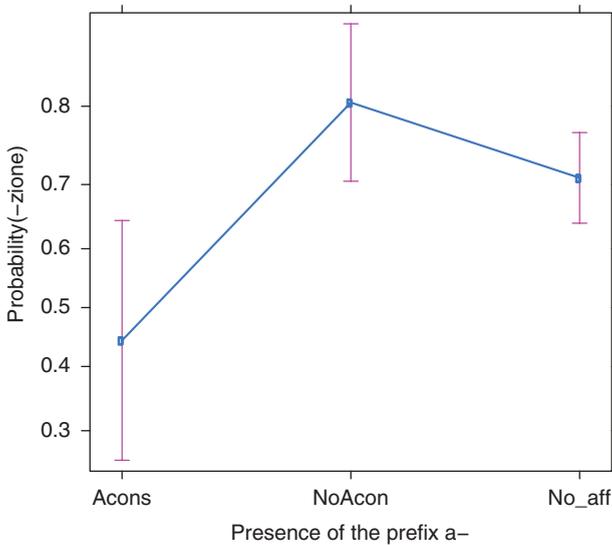
Fig. 4: Effect of the prefix a-.

6. Discussion

The present study contributes to the understanding of the intricate matter of the rivalry between two Italian event-denoting suffixes. The study has focused on the possible constraints that shape the domain of application, by investigating seven possible features. The multivariate statistical analysis has been shown to be useful in the evaluation of the contribution of the different features and it has only partially confirmed previous claims on the topic.

Specifically, the analysis confirms an association between prefixed verbs in *a-* and *-mento* nominalizations, as well as a relation between this suffix and verbs from the third conjugation. By considering these findings together, we can draw two considerations.

First, most of *a-* prefixed verbs (43 out of 51) belong to the first inflectional class. This is an interesting fact if we consider that our analysis also reveals an association between *-zione* EDNs and the first conjugation (and on the other side, an association between *-mento* and the second and third conjugations). This means that even if *-zione* derivatives show higher probabilities to be formed from verbs of the first conjugation, *-mento* nominalizations undermine this association when the verb shows the prefix *a-*.

Second, previous work stated that there is a relation between *-mento* and verbs from the third conjugation that were formed by parasynthesis specifically. Our analysis reveals an association between *-mento* and the whole category of third conjugation verbs, although a qualitative inspection indicates

not only that the majority of third conjugation verbs that form a *-mento* EDN are parasynthetic verbs, but also that prefixed verbs in *a-* from the first conjugation (forming *-mento* derivatives) involve parasynthesis. It is thus possible indeed that parasynthesis is the underlying constraint that drives the choice of *-mento* EDNs. Further work, on a new sample of data, should test this hypothesis by considering parasynthesis as an independent and codified factor in the analysis.

We also observed a relation between verbs from the second conjugation and the *-mento* suffix. The presence of parasynthetic verbs is marginal here, but further work should better assess whether parasynthesis or other factors specific to the conjugation may also play a role in this relation.

With regard to the presence of other affixes, the prefixes *de-, in-* and *s-* were not significant in our study, contrary to what was claimed in the past literature (e.g. Gaeta 2004). The contribution of some affixes could not be assessed due to the small sample and the presence of sampling zeros. However, simple type frequencies (Tab. 7) show that verbs ending in *-ificare* and *-izzare* form only *-zione* derivatives, as claimed in past works. The statistical significance of this association should be assessed on the basis of a larger sample. Indeed, considering the verbs in *-ificare*, we observe that they could form *-mento* EDNs in principle. For example, *purificamento* (from *purificare*, "to purify") is not attested in our sample but is listed in the Treccani online dictionary[12] and it counts around 1400 results in a google search. It is obviously less frequent than its *-zione* counterpart (*purificazione)*, which produces more than 2 million results on google, but further work should test this statistically.

The analysis reveals an association between the total number of morphological processes the derivative has undergone and the EDN suffix: *-zione* nominalizations are associated with fewer morphological processes, whereas *-mento* ones to a higher number. This result contradicts previous claims that mention a slight association between *-mento* and simple base verbs but is in line with what we have seen about parasynthesis: *-mento* derivatives tend to be related to parasynthetic verbs and, since parasynthesis is counted as two morphological processes (i.e. the prefix and the conversion process), it can be responsible for the higher number of processes in *-mento* derivatives. Interestingly, the effect of the number of morphological processes and the effect of verb length go in the opposite direction. Nominalizations in *-zione* have longer bases, even if they show fewer morphological processes.

---

12 http://www.treccani.it/vocabolario/purificamento/

Obviously, there are many other possible features that may constrain the productivity of the two nominalizations considered. I hope to address some in future work.s

7. Conclusion

In the present work, I investigated the role of base constraints in the pattern selection for event-denoting nominalizations. I focused on two suffixes available in Italian, namely *-mento* and *-zione,* and carried a statistical analysis on corpus data. I found that different morphological features of the base verbs (specifically, the length in characters, the inflectional class, and the presence and number of other affixations) influence the use of one pattern over the other. Nominalizations in *-zione* are associated with verbs from the first conjugation, but not if the base verb is prefixed in *a-*. In these cases, *-mento* is preferred. Nominalizations in *-mento* are also preferred for verbs from the second and third conjugations. Moreover, *-zione* base verbs are longer than those forming *-mento* EDN, even if *-mento* derivatives show a higher number of morphological processes. I hypothesize that all these findings may be linked to verbs formed by means of parasynthesis, but future work should test this interpretation.

The results here observed are only a first step towards solving this intricate puzzle, and further research is needed to overcome the limitations of this study. In future work, I intend to expand on the sample coverage, by using a larger corpus, and to employ further factors in the modelling, which may be relevant for this case of affix rivalry. Moreover, I would like to include a diachronic perspective, comparing results from different time periods, and to conduct an additional analysis restricted to neologisms[13], in order to evaluate constraints on *potential* rather than *realized* productivity. Lastly, even if the two suffixes considered are the most productive, future research should take into account other EDN patterns as well.

At a general level, the present study shows that productivity constraints do not represent strict rules with binary outcomes, but they rather emerge as preferences with a graded effect. For instance, stating that verbs from the second conjugation are more frequently associated with the suffix *-mento* is

---

13 A first attempt in this sense was already conducted: the original dataset was restricted considering only EDNs occurring once. Previous studies (e.g. Baayen and Renouf 1996, Plag 2003) suggested indeed that hapax legomena are a good approximation of neologisms. Results of the regression model on these data are in line with those reported above, with the difference of the disappearance of the effect for the length of the base verb. I do not report this analysis here because of the scarsity of data: there were indeed only 120 observations left.

different from arguing that they necessarily select this suffix. Indeed, many cases which would contradict such rigid arguments can be found even in our small sample. Discovering tendencies should not be seen as having a minor impact in research, since they enrich our comprehension of language and language processing. Word formation is a complex process and a complex explanation is what we would expect.

Acknowledgments

References

Adams, Valerie. 1973. *An Introduction to Modern English Word-Formation*. London: Longman.

Agresti, Alan. 2003. *Categorical data analysis*. Hoboken, NJ: John Wiley & Sons.

Arndt-Lappe, Sabine. 2014. Analogy in suffix rivalry: The case of English -ity and -ness. *English Language & Linguistics* 18(3). 497–548.

Baayen, R. Harald & Antoinette Renouf. 1996. Chronicling The Times: Productive lexical innovations in an English newspaper. *Language* 72. 69–96.

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

Bach, Emmon. 1986. The algebra of events. *Linguistics and Philosophy* 9(1). 5–16.

Bauer, Laurie. 2001. *Morphological Productivity*. Cambridge: Cambridge University Press.

Bauer, Laurie. 2005. Productivity: theories. In Štekauer, Pavol & Rochelle Lieber (eds.), *Handbook of word-formation*, 315–334. Dordrecht: Springer.

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

Booij, Geert & Rochelle Lieber. 2004. On the paradigmatic nature of affixal semantics in English and Dutch. *Linguistics* 42(2). 327–357.

Bonami, Olivier, & Juliette Thuilier. 2019. A statistical approach to rivalry in lexeme formation: French-iser and-ifier. *Word Structure* 12(1). 4–41.

Comrie, Bernard & Sandra A. Thompson. 2007. Lexical nominalization. In Timothy Shopen (ed.), *Language Typology and Syntactic Description vol. III: Grammatical Categories and the Lexicon*, 334–381. Cambridge: Cambridge University Press.

Comrie, Bernard. 1976. The syntax of action nominals: A cross-language study. *Lingua* 40(2). 177–201.

Fiorentino, Giuliana. 2008. Action nouns and the infinitive in Italian. *Romanische Forschungen* 120(3). 3–28.

Fox, John, Sanford Weisberg, Michael Friendly, Jangman Hong, Robert Andersen, David Firth & Steve Taylor. 2019. "*Package 'effects'*." ftp://gnualpha.uib.no/pub/cran/web/packages/effects/effects.pdf (accessed 1 February 2020).

Gaeta, Livio. 2002. *Quando i verbi compaiono come nomi. Un saggio di Morfologia Naturale*. Milano: Franco Angeli.

Gaeta, Livio. 2004. Derivazione nominale deverbale. In Maria Grossmann & Franz Rainer (eds.), *La formazione delle parole in italiano*, 314–351. Tübingen: Niemeyer.

Gaeta, Livio. 2015. Restrictions in word-formation. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.*), Word-Formation. An International Handbook of the Languages of Europe*, Berlin & New York: de Gruyter.

Gaeta, Livio. 2017. Nomi deverbali in MIDIA. In Paolo D'Achille & Maria Grossman (eds.*), Per la storia della formazione delle parole in italiano. Un nuovo corpus in rete (MIDIA) e nuove prospettive di studio*, 221–241. Milano: Franco Cesati.

Gaeta, Livio & Davide Ricca. 2002. Corpora testuali e produttività morfologica: i nomi d'azione italiani in due annate della Stampa (1996-1997). In Roland Bauer & Hans Goebl (eds.), *Parallela IX. Testo-variazione- informatica/Text-Variation-Informatik. Atti del IX Incontro italoaustriaco dei linguisti, Salzburg, 1-4 novembre 2000*, 223–249. Wilhelmsfeld: Egert.

Hay, Jennifer. 2001. Lexical frequency in morphology: Is everything relative? *Linguistics* 39(4), 1041–1070.

Iacobini, Claudio & Anna Maria Thornton. 1992. Tendenze nella formazione delle parole nell'italiano del ventesimo secolo. In Bruno Moretti, Dario Petrini & Sandro Bianconi (eds.), *Linee di tendenza nell'italiano contemporaneo. Atti del XXV Congresso della Società di Linguistica Italiana (SLI), Lugano, 19–21 settembre 1991*, 25–55. Roma: Bulzoni.

Koptjevskaja-Tamm, Maria. 1993. *Nominalizations*. London: Routledge.

Melloni, Chiara. 2007. *Polysemy in word formation: the case of deverbal nominals*. PhD Thesis. Verona: University of Verona.

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

R Core Team. 2015. *R: A language and environment for statistical computing*.

Rainer, Franz. 2005. Constraints on productivity. In Pavol Štekauer & Rochelle Lieber (eds.), *Handbook of word-formation*, 335–352. Dordrecht: Springer.

Rainer, Franz, Francesco Gardani, Wolfgang U. Dressler & Hans C. Luschützky (eds.). 2019. *Competition in Inflection and Word-formation*. Cham: Springer International Publishing.

Scalise, Sergio. 1983. *Morfologia lessicale*. Padova: Clesp.

Štichauer, Pavel. 2009. Morphological productivity in diachrony: The case of the deverbal nouns in *-mento*, *-zione* and *-gione* in Old Italian from the 13th to the 16th century. In Fabio Montermini, Gilles Boyé & Jesse Tseng (eds.), *Selected Proceedings of the 6th Décembrettes*, 138–147. Somerville, MA: Cascadilla Proceedings Project.

Talamo, Luigi, Chiara Celata & Pier Marco Bertinetto. 2016. DerIvaTario: An annotated lexicon of Italian derivatives. *Word Structure* 9(1), 72–102.

Thornton, Anna Maria. 1988. *Sui nomina actionis in italiano*. Phd Thesis. Pisa: University of Pisa.

Thornton, Anna Maria. 1990. Sui deverbali italiani in-mento e-zione (I). *Archivio glottologico italiano* 75(2). 169–207.

Thornton, Anna Maria. 1991. Sui deverbali italiani in-mento e-zione (II). *Archivio Glottologico Italiano* 76(1). 79–102.

Thornton, Anna Maria. 2015. Paradigmatically determined allomorphy: the "participial stem" from Latin to Italian. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-Formation. An International Handbook of the Languages of Europe*, 780–802. Berlin & Boston: de Gruyter.

Varvara, Rossella. 2017. *Verbs as nouns: Empirical investigations on event-denoting nominalizations*. PhD Thesis. Trento: University of Trento.

Varvara, Rossella. 2019. Misurare la produttività morfologica: i nomi d'azione nell'italiano del ventunesimo secolo. In Bruno Moretti, Aline Kunz, Silvia Natale & Etna Krakenberger (eds.), *Le tendenze dell'italiano contemporaneo rivisitate. Atti del LII Congresso Internazionale di Studi della Società di Linguistica Italiana (Berna, 6–8 settembre 2018)*, 187–201. Milano: Officinaventuno

Vendler, Z. 1957. Verbs and times. *The Philosophical Review* 66(2), 143–160.

Rossella Varvara
Università degli Studi di Firenze
DILEF
via della Pergola 58–60
50121 Firenze, Italia
rossella.varvara@unifi.it