

Using distributional data to explore derivational undermarkedness: a study of the event/property polysemy in nominalization

Fabio Montermini

CLLE-ERSS

CNRS & Université de Toulouse 2 Jean Jaurès

fabio.montermini@univ-tlse2.fr

Abstract

This paper proposes a corpus-based analysis of deverbal suffixed nouns in Italian displaying an ambiguity between a clear event reading (*partenza* ‘departure’) and a clear property reading (*intelligenza* ‘intelligence’). It focuses, in particular on words derived with the suffixes *-nza* and *-zione*. Three sets of syntactic contexts for words containing the two suffixes (high- and low-frequency *-nza* words and high-frequency *-zione* words) were extracted from a large corpus of contemporary Italian and coded according to their semantic reading. The comparison of the three datasets, on the one side, confirms an evolution, already observed in the literature, of *-nza* from a typically deverbal action suffix to a typically deadjectival property suffix, and, on the other side, shows that the same ambiguity is observed with *-zione* nouns, although, unlike the case of *-nza*, in this case it remains a marginal feature. The results obtained show the interest of large-scale empirical observations for the analysis of morphological phenomena, and militate in favour of a model in which (regular) polysemy should be considered as a constituting property of derived words.

1 Introduction

Although having a strong empirical basis is an important feature of most current studies of morphological derivational phenomena, these are often realized on (sometimes very large) series of complex words taken in isolation, or on the basis of some examples which are intended to exemplify the totality of the uses a derived lexeme can enter into, or at least the most common, ‘unmarked’, ones. This approach is reductive, how-

ever, especially in the study of the semantics of derivational processes, given the pervasiveness and systematicity of such phenomena as polysemy, semantic underspecification, etc. The first goal of this talk is thus to present arguments in favor of an usage-based model of derivational morphology, i.e. an approach in which the properties of complex lexemes, and the rules by which they are formed, are investigated via a thorough observation of their real contexts of use. The perspective adopted here is an exemplar-based one, in the sense that morphological competence is considered to emerge on the basis of the linguistic material speakers are exposed to, and that this dynamics can be simulated by taking into account large amounts of real usage data. The analysis presented can also be qualified as distributional, since it is inspired, in its fundamental assumption, by distributionalist approaches which are current in semantics (cf. Lenci, 2008 for an overview), according to which there is a correlation between a unit’s meaning and its syntactic distribution.

The second goal of the talk is to provide evidence in favor of a non-compositional view of morphological derivation, according to which the semantic properties of complex lexemes cannot be simply computed on the basis of the sub-elements they contain, but rather on the basis of the lexical relations they enter into. The lack of full isomorphism between the form and the meaning of complex lexemes has been observed and investigated in many cases and in many languages. These include cases of over-marking, where an element (e.g. an affix) is present without carrying any evident meaning (cf. Roché, 2009, among others, for several examples in French), and parallel cases of under-marking, where a relevant semantic differentiation lacks an overt formal counterpart. The existence of the latter has been observed since a long time, and is linked with several other phenomena which are well known in the literature on morphology and

Copyright © by Fabio Montermini. Copying permitted for private and academic purposes.

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at <http://ceur-ws.org>

semantics, including (regular) polysemy (Booij, 2010: 78-79), multifunctionality (Luschützky and Rainer, 2013), morphological recycling (Hathout, 2011), etc. Morphological under-marking, precisely, constitutes the main focus of the analysis proposed here.

2 The data

In several languages, deverbal nouns present several instances of systematic polysemy, some of which are well described in the literature (e.g. action / result, cf. Rainer, 1996, Bisetto and Melloni, 2007). In particular, this paper is focused on cases of nominalization which, in spite of their frequency, have received less attention (but cf. Kerleroux, 2008 on French) namely deverbal nouns displaying an ambiguity between an event and a property reading, as in the following examples for the lexeme *vigilanza* in Italian:

(1)
la polizia ha effettuato una vigilanza continua
'police guaranteed a continuous control'

vs.

la sua vigilanza è calata del 50%
'his/her attention decreased of 50%'

Although Italian is the main focus of this paper, it should be observed that the same ambiguity can be observed in other Romance languages (and in English), involving several cognate affixes, such as those derived from Latin *-antia*, *-atio*, *-mentum*, *-tura*. In fact, this ambiguity should probably be ascribed to a specific property deverbal suffixes possessed in Latin (cf. (2)), since it is not observed with other morphological processes which cannot be directly linked to corresponding Latin constructions, such as verb-noun conversions or the (Germanic) deverbal suffix *-al* in English:

(2)
Lat.: *adaequatio* ('adequacy'), *observantia* ('observation')

The polysemy in question can also be linked to the larger spectrum of meanings that have been observed for deverbal nouns; the typical event reading and the typical property reading, in fact, can be considered as the two poles of a continuum which includes the nominalization of more or less permanent states (cf. Fradin, 2011, 2014):

(3)
una partenza / *latitanza / *intelligenza istantanea
'an instant departure / lam / intelligence'

una *partenza / latitanza / *intelligenza di due mesi
'a two-month departure / lam / intelligence'

una *partenza / *latitanza / intelligenza ammirevole
'an admirable departure / lam / intelligence'

Roughly, we can distinguish the three types above according to four dimensions, as exemplified in Table 1.

	action	punctual	bound	quantifiable
partenza	+	+	+	+
latitanza	+	-	+	-
intelligenza	-	-	-	-

Table 1: Types of deverbal nouns.

More specifically, the analysis presented has been carried on on nominalizations containing the two suffixes *-nza* and *-zione*¹, which share the property that, when they are constructed on a verb, they are linked, formally and semantically, to its participle (respectively, the present and the past participles) or to the homophonous adjective (*accogliere* / *accogliente* ⇒ *accoglienza* 'acceptance'; *educare* / *educato* ⇒ *educazione* 'education'). In addition, they can also be constructed on an adjective lacking a verbal counterpart (cf. *frequente* ⇒ *frequenza* 'frequency'; *perfetto* ⇒ *perfezione* 'perfection'), and in this case, base adjectives more often correspond to an individual-level predicate. In spite of their similarities, however, derived nouns in *-nza* and in *-zione* present several important differences. The most relevant one is probably the fact that while *-nza* is mainly attached to stative verbs (cf. Gaeta 2002), i.e. verbs which are semantically closer to (individual-level) adjectives (cf. Chierchia 1995: 177), no such tendency is observed with *-zione*, which, on the contrary, seems to display a preference for active event verbs. Consequently, apart from some exceptions (cf. *partenza* 'departure'), the property reading can be virtually applied to all *-nza* nouns, at least in some of their uses, while for *-zione* the situation is reversed:

¹ In fact, both suffixes may present several different forms in surface, whose selection depends on the form of the base they attach to. The forms given are intended to be labels for more abstract formal representations (on the formal problems posed by *-nza* and *-zione* cf., respectively, Gaeta, 2002: 127-129; Gaeta 2004, 346-348; Thornton, 1990; Montermini, 2010).

most of them do not allow this reading, while others accept it, a behavior for which no clear systematicity can be identified:

- (4)
- a. *determinazione*
 ‘determination’ / ‘determinedness’
- educazione*
 ‘education’ / ‘educatedness’
- b. *istruzione*
 ‘instruction’ / *‘educatedness’
- risoluzione*
 ‘resolution’ / *‘determinedness’

In addition to the general features described above, some empirical observations motivate a deeper large-scale observation of the two derivational processes in question. First, for some of the *-nza* nouns displaying an event reading there exists a corresponding noun containing extra morphological material denoting a property (cf. *assistenza* ‘assistance’ ⇒ *assistenzialità*); similarly, to a past participle can correspond a derived noun denoting a property, either in concurrence with a *-zione* noun or not (cf. *risolto* ⇒ *risolutezza*, *determinato* ⇒ *determinatezza* (vs. *determinazione*). Second, the observation of real language use shows that lexemes with a typical event meaning can be used as property nouns, and vice-versa, like in the following examples taken from the Web:

- (5) La produzione basata sulla **concorrenza** del prezzo tende a tagliare i costi sostenuti dalla produzione basata sulla qualità.
 ‘Production based on low prices (lit. price concurrence) tends to cut the costs incurred by quality based production’.

Paolo [...] era un uomo di estrazione nobile, di grande educazione e **istruzione** ed estremamente religioso e timorato di Dio.
 ‘St. Paul [...] was a man of noble lineage, highly educated and very cultivated (lit. of great education and instruction), and extremely religious and God-fearing’.

Finally, as shown by Benincà and Penello (2005), and as confirmed by the data I have analyzed, while nouns with a pure event reading were the predominant output for *-nza* in ancient Italian, it is more employed today for the con-

struction of property nouns. On the other hand, no comparable shift can be observed for *-zione*.

3 The analysis

In order to test the distribution of meanings for *-nza* and *-zione* nouns, in particular along the event / property divide, I extracted the 61 most frequent lexemes in *-anza* and *-enza* (the two possible formal variants)² in the CorIs³, a large corpus of written Italian. For each of the lexemes in question, 100 contexts of occurrence were randomly selected, each of which was semantically coded according to its compatibility with one of the two meanings in question. In particular, the coding was based on such properties as the possibility of being determined by quantification or a measure adjective, or the presence / absence of temporal boundaries.

Figure 1 shows the distribution of meanings according to the class of the base (verb vs. adjective), and, as expected, a strong correlation between verbal bases and event reading, on the one side, and adjectival bases and property reading, on the other, are observed. The diagram also shows that, for the most frequent *-nza* nouns, the two schemes are more or less equally available.

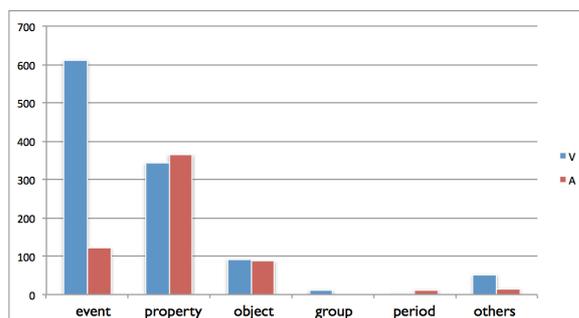


Figure 1: Distribution of meanings for the most frequent *-nza* nouns according to the class of the base.

In order to measure the functioning of this morphological process in the speakers’ synchronic competence, the same procedure was applied to low-frequency words containing *-nza* in the same corpus (62 lexemes overall having a frequency ≤ 3, 88 contexts overall). In this case, as shown in Figure 2, a strong preference of *-nza* for adjectival bases and for the property reading can be observed.

² The lexemes in question range from *presenza* (frequency 19,671) to *indifferenza* (frequency 1,671). Of course, the least was cleared from all lexemes ending in *-nza* that could not be clearly linked to a synchronically existing word.

³ http://corpora.dslo.unibo.it/coris_ita.html.

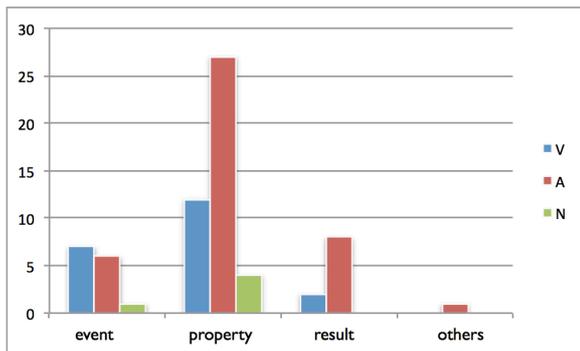


Figure 2: Distribution of meanings for the low-frequency *-nza* nouns according to the class of the base.

This result confirms the observation, mentioned above, that in the history of Italian *-nza* evolved from a (mainly) deverbal suffix forming event nouns to a (mainly) deadjectival suffix forming property nouns.

Moreover, in order to determine whether the features identified for *-nza* are specific to this construction or belong to deverbal nominal suffixes in general, the same procedure was further applied to a comparable set of words in *-zione* (including some possible formal variants, like *-sione*), namely the 61 most frequent forms in the CorIs⁴.

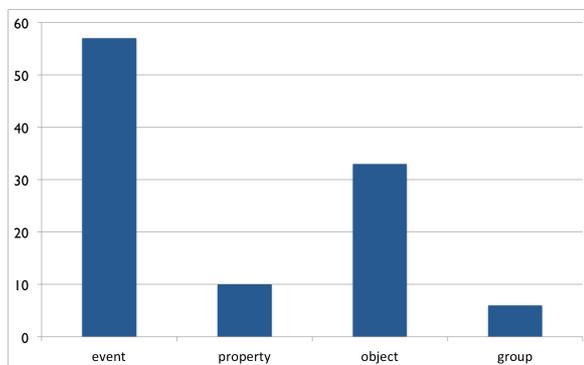


Figure 3: Distribution of meanings for *-zione* nouns.

As Figure 3 shows, the event / property polysemy remains marginal for *-zione*, thus suggesting that, while this polysemy can be considered as a constitutive property of the *-nza* word formation pattern, while it constitutes a rare and marked subpattern for *-zione*.

⁴ Ranging from *amministrazione* (frequency 17,139) to *previsione* (frequency 3,987).

References

- Paola Benincà, and Nicoletta Penello. 2005. Il suffisso *-anza/-enza* tra sincronia e diacronia. In M. Grossmann and A.M. Thornton (eds.), *La formazione delle parole. Atti del XXXVII Congresso internazionale di studi della Società di linguistica italiana. L'Aquila, 25-27 settembre 2003*, 69-86. Roma: Bulzoni.
- Antonietta Bisetto, and Chiara Melloni. 2007. Result nominals: A lexical-semantic investigation. In G. Booij, B. Fradin, E. Guevara, S. Scalise and A. Ralli (eds.), *On-line Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5). Fréjus, 15-18 September 2005*. Bologna: Università di Bologna.
- Geert Booij. 2010. *Construction Morphology*. Oxford University Press, Oxford.
- Gennario Chierchia. 1995. Individual-level predicates as inherent generics. In: G.N. Carlson and F.J. Pelletier (eds.) *The Generic Book*, 176-223. Chicago / London: University of Chicago Press.
- Bernard Fradin. 2011. Remarks on state denoting nominalizations. *Recherches linguistiques de Vincennes*, 40: 73-99.
- Bernard Fradin. 2014. La variante et le double. In: F. Villoing, S. David and S. Leroy (eds.) *Foisonnements morphologiques. Études en hommage à Françoise Kerleroux*, 109-147. Nanterre: Presses Universitaires de Paris Ouest.
- Livio Gaeta. 2002. *Quando i verbi compaiono come nomi. Un saggio di morfologia naturale*. FrancoAngeli, Milano.
- Nabil Hathout. 2011. Une approche typologique de la construction des mots. In: M. Roché, G. Boyé, N. Hathout, S. Lignon, M. Plénat, *Des unités morphologiques au lexique*, 251-318. Paris: Hermès – Lavoisier.
- Françoise Kerleroux. 2008. Des noms indisctincts. In: B. Fradin (ed.), *La raison morphologique. Hommage à la mémoire de Danielle Corbin*, 113-132. Amsterdam / Philadelphia: John Benjamins.
- Alessandro Lenci. 2008. Distributional semantics in linguistic and cognitive research. A foreword. *Rivista di Linguistica*, 20.1: 1-30.
- Hans C. Luschützky, and Franz Rainer. 2013. Instrument and place nouns: A typological and diachronic perspective. *Linguistics*, 51.6: 1301-1359.
- Fabio Montermini. 2010. The lexical representation of nouns and adjectives in Romance languages. *Recherches linguistiques de Vincennes*, 39: 135-161.

Franz Rainer. 1996. La polysémie des noms abstraits: historique et état de la question. In: N. Flaux, M. Glatigny, and M. Samain (eds.), *Les noms abstraits. Histoire et théorie. Actes du colloque de Dunkerque (15-18 septembre 1992)*, 116-126. Villeneuve d'Ascq: Presses Universitaires du Septentrion.

Michel Roché. 2009. Pour une morphologie *lexicale*. *Mémoires de la Société de Linguistique de Paris. Nouvelle série*, 17: 65-87.

Anna M. Thornton 1990/1991. Sui deverbali italiani in *-mento* e *-zione*. *Archivio glottologico italiano*, LXXV (2): 169-207 / LXXVI (1): 79-102.