

Proper Names and Polysemy: from a Lexicographic Experience

Rita Marinelli

Istituto di Linguistica Computazionale C.N.R., Area della Ricerca, Via Moruzzi 1 56124 Pisa Italy

e-mail: Rita.Marinelli@ilc.cnr.it

Abstract

In the framework of the SI-TAL (Integrated Systems for the Automatic Treatment of Language) project the lexical coverage of IWN has been extended by adding, besides two grammatical categories not encoded in EWN (i.e. adjectives and adverbs), a set of proper names which are taken into consideration in this paper. This decision was also due to the high degree of incidence of proper names observed in the corpus selected within SI-TAL for semantic annotation.

In this paper we would refer more widely about the relations involving the **pn** in particular codifying the relation between the **pn** and the senses (literal, derived and extended). We consider the **pn** as the basis for many extensions of meaning.

In fact, many types of derivatives and sense extensions are generated, by means of lexical rules that operate as “generative factors”.

Novel usages of a word form can be derived through productive application of a lexical rule; therefore we propose to represent these lexical rules codifying new semantic relations in the database.

We want to give prominence to the polysemy of **pn** to confirm the linguistic manifestation(s) of the faculty for generative categorization and compositional thought ” (Pustejovsky, 2001).

1. Introduction

IWN (ItalWordNet) has been built enlarging the Italian WordNet developed in the framework of the European project EWN (EuroWordNet) by codifying new grammatical categories (adjectives and adverbs) and a subset of proper names (**pn**).

We aim at focusing this subset, mainly to achieve a well reasoned and structured enlarging of the database, also through the deeper study of the semantic relations involving the **pn**, on the basis of the recent experience carried out in IWN.

Within IWN, the ‘synonymy’ relation applies to the variants of a synset allowing to interchange the synonyms (or variants) in at least one proposition, and this kind of relation is valid also for the set of proper names as formalized below:

$$a = b \Leftrightarrow \{a.f(a)\} = \{b.f(b)\}$$

In IWN, the relation ‘*belongs to class*’ and its reversed ‘*has instance*’ connect ‘instances’ with ‘synsets’: only ‘inherence’ propositions (individual-class) are applicable to **pn** and not ‘relation’ propositions (class-class); the hyponymy relation ‘*is a*’ is not applicable to **pn**.

The subset of **pn** consists, up to now, of more than 4000 **pn**, originated from a first subset of geographic **pn**, further increased with data coming from sources of several type: atlases, Web sites, lists of various kind.

More than 200 classes of **pn** were defined. The database was also enriched encoding other relations involving **pn**.

In fact, it was evidenced that many **pn** are the basis of many substantives and adjectives as their derivatives (e.g.: Nicot-nicotine) and that, when an adjective derived from the **pn** does not exist, **pn** are used very often in appositive/attributive position, e.g.: Braille alphabet. (Marinelli and Roventini, 2002).

The *pertains_to* relation and its reverse *has_pertained*, has been used both in WN and in EWN. It allows the link of a noun with a relational adjective.

In IWN this relation applies either between synsets or between synsets and instances: it connects 2nd order entities with 1st order entities, or 2nd order entities and instances:

dantesco (dantean)	<i>pertains_to</i>	Dante
musicale (musical)	<i>pertains_to</i>	musica (music)

Like the other grammatical categories, also **pn** were linked with WordNet 1.5 by means of equivalence relations. The *eq_synonymy* is used to map proper names with an equivalent instance in WN; the *eq_belongs_to_class*, that was not present in EWN, is codified in IWN to link proper names to the generic belonging class when they have no equivalent in WordNet.

In the following examples all the types of relations so far encoded for this subset are shown:

Roma	<i>belongs_to_class</i>	città (city, town)
Romano	<i>pertains_to</i>	Roma (Rome)
Roma	<i>derivation</i>	romanità (Roman world)
Roma	<i>eq_synonymy</i>	Rome
Livorno	<i>eq_belongs_to_class</i>	town

2. Polysemy of Proper Names

In this paper we would try to refer more widely about the relations involving the **pn**, in particular codifying the relation between the **pn** and the senses (literal, derived and extended). Regular polysemy has been widely studied, closely connected with linguistic phenomena such as metonymy.

Our purpose is to highlight particular cases of polysemy found in codifying **pn**.

The study of polysemy in **pn** may, in our advice, constitute the starting point for a more general ‘theory of

polysemy' concerning also the other grammatical categories. To study the sense shifting mechanisms in this subset is helpful for understanding and describing more sophisticated processes of transposition that are enriching the written and spoken language in every day life.

We consider **pn** as the basis for many extensions of meaning: this may happen when "a more general human metarepresentational capacity" is exploited (Papafragou, 1995). In fact, many types of derivatives and sense extensions are generated, by means of lexical rules that operate as "generative factors" (Pustejovsky, 1995). Novel usages of a word form can be derived through productive application of a lexical rule; therefore we want to represent these lexical rules codifying new semantic relations in the database.

A polysemic production happens by means of a metaphoric use of the **pn** (Your husband is a Croesus), by means of metonymy (to read Dante), or by means of lexical mechanisms like analogy/synecdoche (he would like to drink a Bloody Mary).

Polysemy can be understood as the result of generative mechanisms. So we want to give prominence to the polysemy of **pn** to confirm the linguistic manifestation(s) of the faculty for generative categorization and compositional thought (Pustejovsky, 2001), that "projective transformations" are applied and that "sense extensions are productive processes which require generative lexical mechanisms" (Copestake and Briscoe, 1996).

Considering our experience, it has been noticed that some deviations from the literal reference are present regularly (Nunberg, 1996), when considering some particular belonging classes, sharing regular semantic relationships; hereafter some examples are shown:

- Place/product e.g.: Shiraz, Shantung, Damasco
- Writer/literary work e.g.: I like Manzoni
- Artist/work of art e.g.: a Picasso was stolen
- Craftsman/artifact e.g.: a Stradivari was found in an old chest
- Town/citizens e.g.: Roma has now its new mayor
- Nation/people e.g.: Only the 60% of Italy voted
- Building/person/Institution e.g.: the Quirinale welcomes the princess
- Person/corporation e.g.: Lacoste, Ford, Skoda
- Corporation/product e.g.: the Ferrari won the Formula 1
- University/town e.g.: Bologna is a good law faculty
- Place/battle/defeat or victory e.g.: Waterloo, Caporetto
- Physician/unit of measurement e.g.: Hertz, Baud, Ohm
- Scientist/discovery/medical analysis e.g.: Doppler
- Musician/composition e.g.: Mozart is too difficult for me
- Region/skiing technique e.g.: Telemark
- Athlete's name/technique e.g.: Fosbury
- Name of the grapes/wine e.g.: Sauvignon

Until now only the '*derivation*' relation has been used e.g.:

Ampere1 *belongs to class* fisico (physician)
ampere2 *belongs to class* unità di misura (unit of measurement)

Ampere1 *derivation* ampere2

It connects variants belonging to different PoSs (Parts of Speech) and applies both to the first and to the second order entities as shown in the examples below:

Dalton *derivation* daltonico (daltonic)
Pastorizzare (pasteurize) *derivation* Pasteur

In these cases *derivation* fits well because it is a morphological relation which links the proper name with its derivatives and viceversa. As in EWN, it is used to encode derivation links when no other semantic relation is available.

In the case of metaphor, instead, there is a substitution on the basis of similarity, and, like similitude, sentences like 'he is a true Casanova' are not reversible.

3. Lexicographic Experiences

3.1 WordNet 2.0

Considering a small subset of **pn** (Adonis, Waterloo, Cinderella, Casanova, Peter Pan, Eden, Cashmere, Champagne, etc.) and comparing their presence in WN 1.5 and in WN 2.0, it has been verified that the same concepts are also present in English; that they are far more numerous in WN 2.0 than in WN 1.5, so we can say that they are taken in greater consideration than before; and each **pn** of this subset is present with two (or more) senses, showing the same extension of meaning as in Italian from the literal to the metaphoric or metonymical sense, even if not codified by means of a relation. E.g.:

1. Chimera, Chimaera -- ((Greek mythology) fire-breathing she-monster with a lion's head and a goat's body and a serpent's tail; daughter of Typhon)
2. **chimera**, chimaera -- (a grotesque product of the imagination)
2. 1. **cashmere** -- (a soft fabric made from the wool of the Cashmere goat)
2. **cashmere** -- (the wool of the Kashmir goat)
3. Kashmir, Cashmere, Jammu and Kashmir -- (an area in southwestern Asia whose sovereignty is disputed between Pakistan and India)

3.2 PAROLE Corpus Evidence

Many examples of the sense shifting phenomenon have been evidenced by textual corpora. Particularly rich in metaphorical uses are newspaper articles, which employ an increasingly impressive language to capture the reader's attention. The effect is of surprising discovery of similarity between the two concepts involved in the metaphor. By means of these semantic procedures discourse is given total enrichment, a semantic 'surplus'. Starting from a set of representative samples of proper names, a research has been carried out on the PAROLE (Marinelli et al. 2002) corpus, containing over 20,000,000 occurrences: various types of "sense shifting" (assimilated to typical cases of regular polysemy or regular sense shifting) have been verified. Moreover, for each proper name considered, more transpositions than expected were found: in the case of **pn** like Maratona (Marathon), there has been an extension of the reference from the place of Greece (base sense) to the athletic specialty (the winner of the marathon...) and then to the true metaphorical sense (Marathon of laws); the meaning of Lacoste has been extended from the tennis player René Lacoste (base sense), with the pseudonym 'Crocodile', leading competitor that helped France to win its first Davis cup in 1927, to the sportswear company and then to the garment (to wear a Lacoste).

It has been verified that the frequent use of a proper name in metaphorical sense makes it become a common noun:

Class	Sense Shifting to	PN	Examples carried from Parole Corpus	Examples Translation
Scientist	Phenomenon/Tools	Doppler	' <i>doppler femorale: utilizzando una tecnica messa a punto...</i> '	('femoral doppler :utilizing a technique set up..')
		Wassermann	' <i>lesse il risultato della Wassermann...</i> '	('he read the Wassermann result..')
Scientist	Illness	Parkinson	' <i>quel tremore era Parkinson e non c'era nulla da vergognarsi</i> '	('that tremor was due to Parkinson ..there was nothing to be ashamed for..')
		Alzheimer	' <i>sopportava il marito malato di Alzheimer...</i> '	('she bore the husband suffering from Alzheimer ')
Person/Place	Situation/Main Quality	Dedalo	' <i>assopito nel dedalo di viuzze e casette della povertà contadina</i> '	('sleepy in the daedal of small roads and houses typical of country poverty')
		Babele	' <i>Coloratissimi, nei formati più stravaganti, in una babele di lingue</i> '	('Very colourful in the oddest shapes in a languages babel ')
		Casanova	' <i>era considerato un vero Casanova</i> '	('he was considered a true Casanova ')
Emblematic Word	People represented by	Islam	' <i>L' Islam approva...</i> '	('The Islam approves')
		Cristianesimo	' <i>Il Cristianesimo non riconosce i diritti...</i> '	('The Christianity does not recognize the claims..')
Person who planned/proposed	Prize/Law/Plan	Nobel	' <i>Rubbia ha vinto il Nobel per la fisica</i> '	('Rubbia has won Nobel for physics')
		Pulitzer	' <i>Cristofer è un premio Pulitzer</i> '	('Cristofer is a Pulitzer prize')
Place	Institution	Quirinale	' <i>così, ritualmente, il Quirinale sottolinea come il presidente del Consiglio e i ministri...</i> '	('so, ritually, the Quirinale emphasizes that the Premier and the Ministers...')
		Palazzo Chigi	' <i>Palazzo Chigi intende proseguire la strada...</i> '	('Palazzo Chigi is going to go ahead on the road')
Town	Basket/Football Team	Livorno	' <i>Il Livorno segnò...</i> '	('the Livorno scored..')
		Cagliari	' <i>Scontento anche l'allenatore del Cagliari</i> '	('Unhappy also the Cagliari trainer')
Town	Set of citizens	Siena	' <i>E' famosa l'ospitalità di Siena</i> '	('It is famous the Siena hospitality')
Nation	Inhabitants	Francia	' <i>La Francia ha votato...</i> '	('The France voted..')
Place of production	Product	Gorgonzola	' <i>non capivo come facesse a mangiare il gorgonzola all'alba</i> '	I did not understand how he could eat gorgonzola cheese at dawn..')
		Cognac	' <i>mise il caffè davanti alla ragazza e un bicchiere di cognac vecchio di cinquant'anni</i> '	('he put the coffee in front of the girl and a fifty years old Cognac glass..')
		Cashmere	' <i>due genialoidi targati Biella, Italia, capitale mondiale del cashmere</i> '	('two geniuses typical of Biella, Italy, world capital for cashmere ')
Inventor, builder, creator	Ind. Unit > manufacturer	Goodyear	' <i>un breve test svolto sul circuito di prova Goodyear di Mireval in Francia</i> '	('a short test performed on Goodyear testing circuit at Mireval, France')
		Ferrari	' <i>i tifosi della Ferrari affidano a lui il ruolo del vendicatore</i> '	('the Ferrari fans entrust to him the revenger role')
		Lacoste	' <i>...d'estate veste sempre Lacoste e d'inverno polo in cachemire...</i> '	('in summer he always wears Lacoste and in winter cachemire shirts')
Artisan, artist	Opera, artifact	Tiziano	' <i>ha presentato la sua trionfale sfilata fra i Tiziano e i Caravaggio</i> '	('he is showing his triumphal fashion parade among the Tizianos ' and the Caravaggios ' at Palazzo Doria')
		Stradivari	' <i>possedeva anche uno Stradivari</i> '	('he also owned a Stradivari ')
		Dante	' <i>Carmelo non sa leggere Dante preferiamo Benigni</i> '	Carmelo is not able to read Dante , we prefer Benigni')

Table I – Examples from “Parole” Corpus

A	B	C	D	E	F	G	H	I
Name	Frequency	Texts	Proper use	=D/B %	Extended	=F/B %	Other	=H/B %
Ferrari	938	166	383	40,83 %	218	23,24 %	337	35,93 %
Siena	523	145	488	93,31 %	35	6,69 %	-	-
Quirinale	487	122	141	28,95 %	335	68,79 %	11	2,26 %
Nobel	249	111	12	4,82 %	237	95,18 %	-	-
Maratona	226	95	9	3,98 %	217	96,02 %	-	-
Strasburgo	221	79	171	77,38 %	50	22,62 %	-	-
Palazzo Madama	211	108	30	14,22 %	181	85,78 %	-	-
Champagne	175	87	3	1,71 %	162	92,57 %	10	5,71 %
Picasso	154	59	129	83,77 %	11	7,14 %	14	9,09 %
Caravaggio	75	34	16	21,33 %	54	72,00 %	5	6,67 %
Babele	70	48	40	57,14 %	18	25,71 %	12	17,14 %
Casanova	65	38	12	18,46 %	13	20,00 %	40	61,54 %
Merlin	61	23	12	19,67 %	49	80,33 %	-	-
Chianti	61	27	52	85,25 %	9	14,75 %	-	-
Rambo	59	40	10	16,95 %	33	55,93 %	16	27,12 %
Doppler	56	7	-	-	56	100,00 %	-	-
Caporetto	51	31	30	58,82 %	21	41,18 %	-	-
Parkinson	42	18	25	59,52 %	10	23,81 %	6	14,29 %
Amarcord	33	26	10	30,30 %	23	69,70 %	-	-
Gorgonzola	33	21	8	24,24 %	25	75,76 %	-	-
Dedalo	32	26	2	6,25 %	16	50,00 %	14	43,75 %
Alzheimer	30	16	19	63,33 %	9	30,00 %	2	6,67 %
Mecenate	27	26	2	7,41 %	20	74,07 %	5	18,52 %
Pulitzer	20	18	9	45,00 %	8	40,00 %	3	15,00 %
Stradivari	13	8	8	61,54 %	3	23,08 %	2	15,38 %
Telemark	3	2	-	-	3	100,00 %	-	-
Total	3915	1381	1621	41,40 %	1816	46,39 %	477	12,18 %

Table II - Statistics

as pointed out by Kilgariff (1992), 'words have an indefinite number of potential senses', but there are cases

in which one sense of a systematic relationship is most salient for certain instantiations of it, and, for some word

forms (cicerone, champagne, etc.), the derived sense will have a higher frequency than the base sense and will prevail on it. What is initially not endorsed as a “truthful description of a referent becomes the proper descriptive meaning” and is registered in the lexicon. (Papafragou., 1995). The frequent use of a **pn** in metaphorical sense, in many cases, makes it be lexicalised as common noun (sometimes the capital initial letter of the proper name is lost, sometimes it remains, even in similar contexts).

In some cases a de-properization is found: *‘the Florence of the twenties’*.

Moreover, it has been noticed that, in many cases, also a change of the ontological ‘value’ takes place with a shift of reference: some proper names belonging to the 1st order entities can shift towards 2nd or 3rd order entities, for example the **pn** indicating diseases (Parkinson), physical/engineering methodologies (Diesel), rehabilitative gymnastics methods (Pilates), surgical operating techniques (Milligan Morgan).

4. Our Proposal

When there is a regular shifting from a class to another belonging class (either in the case of metonym or in the case of metaphor), also for **pn** we deemed worthy to find a specific code representing this phenomenon in IWN: we propose to indicate the **pn** regular shifting using this code: *‘has extension’* and its reversed *‘is extension of’* e.g.:

Firenze1	<i>belongs to class</i>	città (town)
Firenze2	<i>belongs to class</i>	cittadinanza, insieme dei cittadini) (municipality)
Firenze1	<i>has extension</i>	Firenze2
Firenze2	<i>is extension of</i>	Firenze1

We think that to make explicit these sense extensions for **pn** could be an useful improvement for the IWN database. Our proposal is to study in detail the various subjects described, which seem worthy of peculiar research; to exploit more widely the other IWN relations. Many more connections will be created that may teach us about mechanisms of metaphor production and comprehension (Fellbaum 2004), considering that ‘the structures underlying the distinct meanings of the words are the heart of the cognitive linguistics enterprise’ (Kilgariff, 1997).

References

- Alonge, A., Calzolari, N., Vossen, P., Bloksma, L., Castellon, I., Marti, T., Peters, W.: The Linguistic Design of the EuroWordNet Database, Special Issue on EuroWordNet, in: N. Ide, D. Greenstein, P. Vossen (eds.), «Computers and the Humanities», XXXII (1998), 2-3, 91-115.
- Copestake A., Briscoe E.J., Semi-productive polysemy and sense extension. Journal of semantics, 12, 15-67, 1995.
- Fellbaum, C. ed.: WordNet: An Electronic Lexical Database, MIT Press, Cambridge, MA, (1998).

- Fellbaum, C. Metaphors in the (mental) lexicon, Proceedings of GWC 2004 Second Global WordNet Conference, Brno, 2004.
- Kilgariff A., Thesis: "Polysemy" University of Sussex, 1992.
- Kilgariff A., I don't believe in word senses, in Computers and the Humanities 31 (2), pp 91-113-University of Brighton ITRI-97-12 July, 1997.
- Marinelli R., Biagini L., Bindi R., Goggi S., Monachini M., Orsolini P., Picchi E., Rossi S., Calzolari N., Zampolli A.: "The Italian Parole Corpus: an Overview". In: Linguistica Computazionale, Giardini Editori, Pisa, pp. 401-422, 2003.
- Marinelli R., Roventini A., Proper Names in a Semantic Database, Proceedings of LREC 2002 Third International Conference on Language Resources and Evaluation, Las Palmas de Gran Canaria, 2002.
- Nunberg G.D., Transfers of meaning, Journal of Semantics, 12, 1995.
- Papafragou A., Figurative language and the semantics-pragmatics distinction, Language and Literature, 5, 179-93, 1996.
- Papafragou A., Metonymy and relevance, UCL Working Papers in Linguistics, 7; 141-175, 1995.
- Peters W., Extraction of implicit Knowledge from WordNet, Proceedings of Ontolex2002 Workshop on Ontologies and Lexical Knowledge Bases, preceding LREC2002, Las Palmas, 2002.
- Pustejovsky J., "The generative lexicon", MIT Press, Cambridge Mass. 1995.
- Pustejovsky J., Generativity and Explanation in Semantics. In: Bouillon P. and Busa F. The Language of Word Meaning, Cambridge University Press, Cambridge, MA, 2001.
- Roventini A., Alonge A., Bertagna F., Calzolari N., Marinelli R., Magnini B., Speranza M. (2002). "ItalWordNet: a Large Semantic Database for the Automatic Treatment of the Italian Language" in: Proceedings of the First Global WordNet Conference, Central Institute of Indian Languages, Mysore, India, pp.1-11.
- Vossen P., Condensed Meaning in EuroWordNet. In: Bouillon P. and Busa F. The Language of Word Meaning, Cambridge University Press, Cambridge, MA, 2001.
- Vossen, P. (ed.): EuroWordNet General Document, 1999. <http://www.hum.uva.nl/~EWN>