Lexical Resources and Ontological Classifications for the Recognition of Proper Names Sense Extension

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

Particular uses of PNs with sense extension are focussed on and inspected taking into account the presence of PNs in lexical semantic databases and electronic corpora.
Methodology to select ad include PNs in semantic databases is described; the use of PNs in corpora of Italian Language is examined and evaluated, analyzing the behaviour of a set of PNs in different periods of time.
Computational resources can facilitate our study in this field in an effective way by helping codify, translate and handle particular cases of polysemy, but also guiding in metaphorical and metonymic sense recognition, supported by the ontological classification of the lexical semantic entities.
The relationship between the "abstract" and the "concrete", which is at the basis of the Conceptual Metaphor perspective, can be considered strictly related to the variation of the ontological values found in our analysis of the PNs and their belonging classes which are codified in the ItalWordNet database.

1. Introduction

ItalWordNet (IWN) is a database of Italian language which was created in the framework of the European project EuroWordNet (EWN). This version of the database was successively updated extending its lexical coverage: in fact, a set of adjectives and adverbs were added, and, besides verbs and nouns that were not present in EWN, also a set of Proper Names (PNs) was included into the Italian wordnet.

The model used to structure the lexicon is based on the central role of the lexical relation of synonymy and on the concept of "synset", namely a set of synonymous words belonging to the same Part-of-Speech (PoS) that can be interchanged in at least one context, according to EuroWordNet (Vossen, 1999) and WordNet (WN) philosophy (Miller, 1990). The database is of relational type. Lexical semantic relations link synsets pair wise: language "internal relations" hold between synsets within the database: i) hierarchical (vertical) relations, instantiated in terms of hyperonymy/hyponymy relationship ("is a"); ii) horizontal relations, such as "part of", "is means for", "role", "belongs to class", etc.

On the base of EWN model, "equivalence relations" link each synset of the Italian wordnet to the closest concepts of the Princeton WordNet (version 1.5).

As far as PNs are concerned, hierarchical relations are not used for their codification: only singular attributive propositions (Blanché, 1968) or "inherence" propositions (individual-class) are applicable to PNs. In such a way each PN is considered as an "instance" of the class which it belongs to: the relation "belongs to class" and its reversed "has instance" connect "instances" with "synsets":

Firenze(Florence) belongs_to_class   *città*
città (town)   has instance   *Firenze*

Other types of internal relations are used to connect PNs to adjectives and nouns:

Fiorentino (florentine) pertains_to   *Firenze*
Firenze (Florence) derivation *fiorentina* (florentine beefsteak)

PNs as well are mapped onto an equivalent instance in WN, through the Inter Lingual Index (ILI)[1], by means of the "eq_synonym" relation; the "eq_belongs_to_class" relation is used to link PNs to the generic belonging class when they have no equivalent synonyms in WN:

Roma   eq_synonym   *Rome*
Livorno eq_belongs_to_class *city,town*

When the closest concept of the PN was not found in the English WN, the PN was included in a list by which the ILI had to be updated and enlarged. Also the set of equivalence relations has to be increased by adding new relations, so that the link with the English can be more fine grained and rich. The wordnet enlarged and strengthened could allow to recognize English PNs that were not found up until now.

By the links to the ILI, PNs are also connected to the Top Ontology (TO) which IWN inherited from EWN, that is a set of concepts with high level of abstraction, hierarchically organized and language independent:

Roma (town) → Artifact, Place
Vespucci (military ship) → Artifact, Object, Vehicle
Aida (opera) → Agentive, Communication, Purpose, Social.

IWN (Roventini et al., 2003) contains almost 50,000 synsets. More than 3800 instances are included in the set of PNs, clustered in about 250 belonging classes.

A first subset of geographic PNs was codified, further increased with data coming from sources of several types: atlases, Web sites, frequency lists obtained from textual

---

[1] The ILI is an unstructured fund of synsets (mainly taken from WorNet1.5), the so-called ILI-records (Vossen, 1998).

corpora, etc. Hereafter some data involving PNs in IWN are shown:

Description	Total
Belongs_to_class	3.858
Has_pertained	350
Derivation	210
Eq_synonym	995
Eq_belongs_to_class	2.462

Table 1. Some data from IWN

In the table 2. the classes of PNs with the greatest number of instances are represented:

Belonging Class	No. of instances
Città (city, town)	568
Porto (harbour)	340
Fiume (river)	275
Teatro (theatre)	273
Museo (museum)	256
Nazione (nation, country)	250
Popolo (citizenry, people)	132
Comune (municipality)	124
Regione (territory, region)	106
Divinità (divinity)	104
Ditta (company)	95
Parco (national park)	80
Cometa (comet)	76
Monte (mountain)	75
Opera lirica (opera)	74
Stella (star)	70
Lago (lake)	68

Table 2. Data about classes and instances

This paper relates about particular cases of metaphorical and metonymic sense extension of PNs which were highlighted while assigning each PN a belonging class, e.g.: "this island is a true Eden", "the Yamaha won". Starting from some examples from the English WN and the many contexts of use found in textual corpora of Italian language, we want to confirm that computational resources and, most of all, the reference to the ontological structure of IWN can contribute to the study and identification of PN sense extension.

2. Lexical Resources and Proper Names

The use of figurative language plays a central role in everyday discourse; the conceptual or cognitive nature of the transfers or mappings which underlie the processes of sense extension has been emphasized by many authors. We follow the perspective of the Conceptual Metaphor Theory whose basic ideas were advanced by George Lakoff and Marc Johnson (Lakoff and Johnson, 1980, see also, among others, Lakoff and Turner, 1989, Lakoff, 1993) which has been highly influential within cognitive linguistics and which posits that metaphor "is serving a

function that cannot be served by literal language" (Katz, 1998), because metaphor "is not a property of language, but rather is a property of our conceptual system" (ibid.); In other words, citing Croft and Cruse (2004), metaphor is the result of a special process for construing a meaning.

Our aim is to examine the many kinds of sense extension starting from the study about the set of PNs present in IWN and their semantic classes.

Corpus-based methods and lexical semantic database approach can be usefully applied to research issues concerning many aspects of word meaning and language use.

The most recent version of the Princeton database, namely WordNet 3.0, was used as reference for examining a set of representative samples of PNs (about 200), chosen on the basis of the frequency lists obtained from the corpora. It was verified that the same concepts (Adonis, Waterloo, Cinderella, Casanova, Peter Pan, Eden, Cashmere, Champagne, etc.) are present in Italian and in English and that each sample of this set is codified in WN 3.0 as a synset with two (or more) senses, and one of them shows an extension of meaning (as in Italian); but only a subset of PNs in WN is coded as "instance"[2], and the sense extension is not represented by means of an "ad hoc" relation:

1. **Casanova**, Giovanni Jacopo Casanova, Giovanni Jacopo Casanova de Seingalt (an Italian adventurer who wrote vivid accounts of his sexual encounters (1725-1798))
   - *instance*
     - (n) adventurer, venturer (a person who enjoys taking risks)
2. **Casanova** (any man noted for his amorous adventures)[3]

It is assumed that observation of large amounts of real and valid data is a viable method for language description (Sinclair, 1991).

Textual corpora can actually show many examples of the sense shifting phenomenon. Particularly rich in metaphorical uses are newspaper articles, which employ an increasingly impressive language to capture the reader's attention.

The set of samples sought and compared in WN 3.0 as well was considered as a starting point to verify the use of PNs in a large corpus of Italian contemporary language, the CLIC corpus *(Corpus di Lingua Italiana Contemporanea)*, consisting of almost 100 million words and constructed after the end of the PAROLE Project (1997) (Marinelli et al., 2003), enlarging the PAROLE Corpus. One of the main goals of this project was to ensure the creation of a comparable set of large Written Language Resources (WLRs) for all the european languages. The CLIC Corpus, enlarged adding data from various types of newspapers, (encoded following the

---

[2] The distinction between hyponyms and classes has been drawn only in recent years (Miller and Hristea, 2006).

[3] From WN web site: http://wordnetweb.princeton.edu/perl/webwn

general standard rules recommended by the 'old' Project), covers the years until 2003.

A diachronic analysis was carried out on two subsets – years 1992/1994 and 1999/2001 - of the CLIC Corpus, constructed on purpose, both containing about 20 million words and the same type of texts (daily newspapers: La Stampa, Repubblica, Il Sole 24 Ore). They were examined and compared using also the lists of concordances provided by the textual database management system (Picchi, 2003) and the sense extensions were manually annotated; see the table hereafter:

```
C:\DBT2000\DBTDATA\DBTPR5.PSB
 ----------------------------------------------------
N.Famiglie definite =  2
N.Famiglie definite =  2  [seq] [1,20]
*casa* & *bianca*
  N. Contesti associati : 3904
Epoca  Mondadori 1985
  1) che vivono in ville lussuose (Castellano addirittura in una che è la copia della
     *Casa *Bianca*), lontani dalla lotta, con l'aria di essere nauseati da
     tutto il baccano che si fa intorno a  - EP15-03-85 LA MAFIA E LE SUE
     CINQUE SORELLE.158.p.30.148
Storia illustrata  Mondadori 1985
  2)Tit3\... \By\di Paolo Maltese \Didasc\  Washington, febbraio 1942: Roosevelt,
     durante una conferenza alla *Casa *Bianca*, dopo l'occupazione
     giapponese di Singapore. - SI00-04-85 GLI MANCO' LA GIOIA  DELLA
     VITTORIA.29.p.55  .17
  3)Tit2\... "L'industria ha responsabilità sociali" \Tit\La *Casa *Bianca* attacca:
     "Aziende senza umanità" DAL NOSTRO- CS05-01-96 La Casa Bianca
     attacca: ".8.p.6-.8
In Versilia e nel tempo  Marco Forti Einaudi 1986
  4)loro, cessarono anche le visite a Lola e alle figlie; ma per anni Giovanni si
```

Table 1. "*Casa Bianca* " (White House) List of concordances - some examples

Many examples of sense extension generated in many sentences of common sense were found. The results showed that the extended senses have a higher frequency than the literal ones and an increasing percentage of extended use of PNs with regular polysemy, in the two different periods of time was verified (Marinelli and Bindi, 2007, Marinelli and Bindi, 2008). Even if the sample sought can be considered too small to argue statistically consistent results, a "gradual fixation" of the meaning in metaphorical/metonymic sense as time passes was evident.

3. Relations in the lexical database

The (theoretical) framework for analyzing the "extended" use of PNs may interact with computational resources and corpus perspective.

Lexical semantic databases have the potential for "translating" in terms of semantic relations the metaphorical processes verified in the many contexts of use of the electronic corpora. Thus, after checking our data in WN 3.0 and verifying them in the corpus, it became evident that some particular belonging classes are candidates for regular shifting of meaning; here are some examples:

− Town/citizens: Roma has now its new mayor
− Person/corporation: Lacoste, Ford, Skoda
− Artist/work of art: a Picasso was stolen
− Place/battle/defeat or victory: Waterloo, Caporetto
− Musician/composition: Mozart is too difficult for me
− Region/skiing technique: Telemark
− Scientist/discovery/medical analysis: Doppler

− Athlete's name/technique: Fosbury

Therefore a semantic relation has been created to represent this phenomenon in IWN (Marinelli 2004): "*has extension*" and its reverse "*is extension of*":

Casa Bianca1   belongs to class   *edificio* (building)
Casa Bianca2   belongs to class   *carica* (office)
Casa Bianca1   has extension   *Casa Bianca2*
Casa Bianca2   is extension of   *Casa Bianca1*

4. Ontological Classification

In Lakoff's words (1993), metaphor is a set of conceptual correspondences, or, more technically, mappings, between two conceptual domains: a source domain (typically more physical or concrete), and a target domain (which is more abstract) that will consist of emotions, behaviour patterns, political strategies, etc. This correspondence is established for the purpose of "understanding the more abstract in terms of the more concrete" (Kovecses, 2000) and makes up a conceptual metaphor.

The correspondence between the "abstract" and the "concrete" has been taken into account focussing on the relationship between the PNs and their belonging classes.

From one hand, the corpus approach gave us the possibility of evaluating the PN-belonging class pairs within the contexts of the sample set from the CLIC corpus. From the other hand, the structure of the semantic database IWN was used as reference for the ontological classification of the nouns which represent the belonging classes. In fact, all the synsets in IWN are ontologically classified according to the Top Ontology (TO); as mentioned above, the concepts of the TO are hierarchically organized and clustered in three main categories or "orders". The noun representing the belonging class can be a concrete noun or an abstract noun. The shifting of PNs from a belonging class to another, in many cases, seemed to correspond to a change of the ontological classification of the class: some belonging classes are ontologically classified as 1st order entities, e.g.: *scrittore* (writer), *edificio* (building), *scienziato* (scientist), etc., others are categorized as 2nd or 3rd order entities, e.g.: *malattia* (desease), *metodologia scientifica* (engineering methodology), *tecnologia software* (software technology), etc.

The ontological conceptual organization of the database (useful in automatic metaphor recognition) can help recognize a sense extension, because a change of the ontological value of the belonging class can be considered diagnostic, as a sign strictly connected with the presence of sense extension.

Let us consider these examples from the CLIC corpus:

− *Spa è un comune del Belgio* (Spa is a municipality of Belgium)

− *Le offerte dell'ultimo minuto sono bagni termali e Spa* (last minute offers are thermal baths and Spa)

The synset *"Spa"* is codified in the database by means of semantic relations:

Spa1 belongs_to_class *comune* (municipality) (1st order entity) → Artifact, Place → literal use of the PN (probably)

Spa2 belongs_to_class *servizio* (service) (2nd order entity) → Agentive, Purpose, Social →sense extension

Considering another example:

–*Pilates ha dedicato sé stesso allo studio della riabilitazione* (Pilates dedicated himself to the study of a rehabilitation technique)

–*Il Pilates rende più tonici i muscoli* (Pilates makes the muscles more tonic)

The synset "Pilates" appears codified in the database in this way:

Pilates1 belongs_to_class *infermiere* (nurse) (1st order entity) →Human, Object →literal use of the PN (probably)

Pilates2 belongs_to_class *tecnica riabilitativa* (rehabilitation technique) (3rd order entity) → Mental, Dynamic → sense extension

Other "novel" examples of sense extension were taken from the Internet to update the set of PNs in IWN, e.g.:

–*L'ANSA raccoglie e trasmette notizie sui principali avvenimenti* (ANSA collects and transmits news about the most important events)

–*E' stato deciso di trasmettere un'ANSA* (it was decided to transmit an ANSA)

The synset "*ANSA*" is codified in the database:

ANSA1 belongs_to_class *agenzia* (news agency) (1st order entity) → Goup, Artifact, Building → literal use of the PN (probably)

ANSA2 belongs_to_class *comunicazione* (communication) (3rd order entity) → Mental, Purpose, Social → sense extension

Another example is:

–*Re Aroldo I di Danimarca (Harold Bluetooth) unì i popoli della penisola scandinava con la religione* (King Harold Bluetooth unified scandinavian peoples by means of religion).

–*Il logo di Bluetooth usa le rune nordiche* ᚼ *(Hagall) e* ᛒ *(Berkanan), analoghe alle nostre H e B* (the Bluetooth logo uses the nordic runes Hagall and Berkanan equivalent to our H and B).

The synset *"Bluetooth"* is so codified in the database:

Bluetooth1 belongs_to_class *re* (king) (1st order entity) → Human, Object →literal use of the PN (probably)

Bluetooth2 belongs_to_class *tecnologia software* (software technology) (3rd order entity) → Purpose, Dynamic → sense extension

If a PN is an instance of two (or more) belonging classes which can be ontologically classified as a 1st order entity (concrete noun) and as a 2nd order or 3rd order entity (abstract noun) respectively, the presence of an abstract noun, entails the presence of a meaning extension, while the presence of a concrete noun, namely a 1st order entity, in the PN-belonging class pair, indicates the (probable) literal use of the PN. We say "probable", not certain, because some examples are found in which a belonging class is a concrete noun showing an extension of meaning: *mangiare Simmental* (to eat Simmental).

The presence of an abstract noun seems to be a sign indicating the use of a sense extension, but our research has to develop new criteria and methodologies to refine methods for the annotation of sense extension in textual corpora and for the analysis of the classes of PNs which are candidate for regular shifting of meaning. In such a way it will be possible to supply more effective instruments for the automatic recognition of sense extension: it will be necessary to use larger sample sets from corpora to be examined and to define new rules and algorithms so as to maximize efficiency for statistical analysis, in such a way as to determine, for each PN, the probability to be used with sense extension.

5. Conclusion

The behaviour of PNs in specific cases of metaphorical and metonymic sense extension has been focussed on and inspected starting from the presence of PNs in lexical semantic databases and electronic corpora.

Methodology to select and include PNs in semantic databases has been described; the use of PNs in corpora of Italian Language has been examined and evaluated, analyzing the contextual use of a set of PNs in different periods of time.

Computational resources can really facilitate our study in this field in an effective way by helping codify, translate and handle particular cases of polysemy, but also guiding in metaphorical and metonymic sense recognition, supported by the ontological classification of the lexical semantic entities.

The relationship between the "abstract" and the "concrete" is at the basis of the Conceptual Metaphor perspective: each conventional metaphor, that is, each mapping across conceptual domains, is a perfectly regular way in which human beings conceptualize and build their non-physical, abstract world (Kovecses, 2005). These correspondences can be considered strictly related to the variation of the ontological values found in our analysis of the PNs and their belonging classes which are codified in the ItalWordNet database.

The abstract-concrete shifting and alternation are reproduced, at database level, in terms of change of the ontological value (concrete-1st order entities –> abstract-

2nd, 3rd order entities), together with the representation of the "abstract" by means of the "concrete"; the established correspondence provides powerful evidence that our conceptual system "organizes" abstract concepts in terms of more concrete kinds of experiences, which help to make the abstract concepts more readily accessible (Evans and Green, 2006).

References

Croft W., Cruse D. A. (2004). *Cognitive Linguistics*. Cambridge University Press.

Evans V., Green M. (2006). *Cognitive Linguistics. An Introduction*. Edinburgh: EUP.

Katz A.N. (1998). Figurative Language and Figurative Thought: a Review. In *Figurative Language and Thought*. Eds. A.N. Katz, C. Cacciari, R.W. Gibbs, M. Turner. New York: Oxford University Press

Kovecses Z. (2000). *Metaphor and emotion: language, culture, and body in human feeling*. New York: Cambridge University Press.

Kovecses Z. 2005. *Metaphor in culture: universality and variation*. Cambridge: Cambridge Universty Press.

Lakoff G. (1993). The Contemporary Theory of Metaphor, In *Metaphor and Thought*. Ed. A. Ortony (2nd edition). Cambridge MA: Cambridge University Press.

Marinelli R., Biagini L., Bindi R., Goggi S., Monachini M., Orsolini P., Picchi E., Rossi S., Calzolari N., Zampolli A. (2003). The Italian Parole Corpus: an Overview. *Linguistica Computazionale*, Pisa, Giardini Editori.

Marinelli R. (2004). Proper Names and Polysemy: From a Lexicographic Experience. *LREC 2004: Lisbon, Portugal, May 2004. Proceedings,* Volume I, Paris, The European Language Resources Association (ELRA).

Marinelli R., Bindi R. (2007). Creativity and Fixation Processes in Proper Names Sense Extensions. In *SILF Actes du XXIXème Colloque International de Linguistique Fonctionnelle*, Helsinki 2005.

Marinelli R., Bindi R. (2008) Uso Metaforico e Metonimico dei Nomi Propri: una Verifica su un Corpus di Italiano Contemporaneo. In (E. Cresti ed.) *Società Internazionale di Linguistica e Filologia Italiana Prospettive nello studio del lessico italiano. Atti del IX Convegno Internazionale SILFI* (Firenze, 14-17 giugno 2006). Firenze, University Press, 2008. Voll.2, I, pp.285-290.

Miller G. A., Beckwith R., Fellbaum C., Gross D., Miller K.J. (1990). Introduction to WordNet: An On-line Lexical Database. *International Journal of Lexicography*.

Miller G.A., Hristea F. (2006). WordNet Nouns: Classes and Instances. *Computational Linguistics*. Volume 32, Issue 1 2006. pp.:1-3.

Sinclair J. M. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.

Vossen P., L. Bloksma, H. Rodriquez, S. Climent, N. Calzolari, A. Roventini, F. Bertagna, A. Alonge, W. Peters. (1998). The EuroWordNet Base Concepts and Top Ontology. *EuroWordNet (LE-4003) Deliverable D017D034D036*, University of Amsterdam.

Vossen, P. (ed.): EuroWordNet General Document, 1999.http://www.hum.uva.nl/~EWN.