The Italian Metaphor Database

Antonietta Alonge

Dipartimento di Filosofia, Linguistica e Letterature – Sezione di Linguistica
University of Perugia
Piazza Morlacchi, 11 – Perugia 06123 – ITALY
anto.alonge@unipg.it

Abstract

This paper describes the main features of the Italian Metaphor Database, being built at the University of Perugia (Italy). The database is being developed as a resource to be used both as a knowledge base on conceptual metaphors in Italian and their lexical expressions, and to enrich general lexical resources. The reason to develop such a database is that most NLP systems have to deal with metaphorical expressions sooner or later but, as previous research has shown, existing lexical resources for Italian do not contain complete and consistent data on metaphors, empirically derived but theoretically motivated. Thus, by referring to the Cognitive Theory of metaphor, conceptual metaphors instantiated in Italian are being represented in the resource, together with data on the way they are expressed in the language (i.e., through lexical units or multiword expressions), examples of them found within a corpus, and data on metaphorical linguistic expressions encoded/missing within ItalWordNet.

1. Introduction

Traditionally, metaphor was considered as a deviant use of language, typically found in genres such as poetry or political oratory. However, around the late 1970s, linguists, psychologists and philosophers started showing its pervasiveness in language, connecting such a phenomenon to cognitive facts. In particular, Lakoff & Johnson’s (1980) seminal work led to a view of metaphor as a central cognitive tool used to structure abstract conceptual domains – the so-called target domains of metaphors – in terms of concrete, and more clearly delineated, domains – the so-called source domains (e.g., the conceptual domains of love, difficulties, and anger in terms of journeys, containers, and heat, respectively).1

Given the pervasiveness of metaphor in language, most NLP systems have to deal with it sooner or later. However, so far very few systems have been designed to handle specifically metaphorical expressions. These are, in general, knowledge-based systems, because they rely on a representation of (a part of) the conventional links (mappings) between metaphorical expressions in the source domains and those in the corresponding target domains. These systems need to acquire a lot of knowledge, thus lexical resources providing rich data on metaphor can be very useful for them (but, of course, also for other kind of NLP systems).

Normally, currently existing general lexical resources do not contain much (explicitly encoded) data on metaphor; furthermore, they do not provide information which can be used to deal with linguistic creativity related with metaphor (i.e., the fact that new metaphorical expressions are often “created”, which people may however understand).

In order to meet the need for extensive data on metaphors, some specific metaphor databases are being built, e.g. for English (for instance, the ATT Meta MetaBank Database - http://www.cs.bham.ac.uk/~jab/ATT-Meta/Databank/) or German/French (the Hamburg Metaphor Database – http://www1.uni-hamburg.de/metapherm/index_en.html). This paper describes a metaphor database being built for Italian at the University of Perugia. The Italian Metaphor Database (IMD) is a resource designed to be used both as a specific repository of information on metaphors in Italian and to enrich general lexical resources. The paper is organized as follows: in section 2 some prominent aspects of the cognitive perspective on metaphor, taken into consideration to design the structure of the IMD, are pointed out; in section 3 issues related with the computational treatment of metaphor and with its representation within linguistic resources are briefly discussed; section 4 is devoted to the description of the structure and content of the IMD; remaining work is then discussed in the Conclusion.

2. Aspects of Cognitive Metaphors

Work in Cognitive Linguistics has shown that “abstract concepts are largely metaphorical” (Lakoff & Johnson, 1999: 3), thus metaphor is pervasive in thought and, consequently, in language, given that metaphorical linguistic expressions are instantiations of conceptual metaphors. Other selected issues raised within the cognitive theory of metaphor are worth mentioning, since they were taken into consideration to outline the structure of the IMD.
Conceptual metaphors arise from sensorimotor experiences (Lakoff & Johnson 1980; Lakoff, 1993; Lakoff & Johnson, 1999), most of which are common to everyone, since they are determined by the way our body is structured, the way we perceive the physical environment, etc. According to Johnson’s theory of conflation (Johnson, 1997), children pass through a “stage of conflation”, during which they do not distinguish between related sensorimotor and non-sensorimotor experiences: for instance, they correlate the subjective experience of affection “with the sensory experience of warmth, the warmth of being held” (Lakoff & Johnson, 1999: 46). There is, then, a period of “differentiation”, during which the conceptual domains are recognized as separate, while the cross-domain association is maintained. These conceptual metaphors emerging out of the two, “conflation” and “differentiation”, stages are so-called “primary metaphors”.

According to Grady (1997), we may distinguish between primary metaphors and complex ones, where the former have a minimal structure and arise through the conflation/differentiation mechanism, and the latter are composed of multiple primary metaphors. Lakoff & Johnson (1999: 56) make clear that we acquire primary metaphors “automatically and unconsciously via the normal process of neural learning [...] When the embodied experiences in the world are universal, then the corresponding primary metaphors are universally acquired”. Thus, universally primary metaphors are inevitably acquired and used, and they are conventionalized in languages. Conventional metaphors form “a huge part of our conceptual system and affect how we think and what we care about” (Lakoff & Johnson, 1999: 60).

A salient property of metaphors, arising as the result of the mapping from a source domain (a concrete or sensorimotor domain) to a target domain (abstract or nonsensorimotor), is that of preserving inference, i.e. inferences which can be made for the source domain also apply to the target domain. A further important corollary of the domains mapping view of metaphor is that existing conceptual metaphors are exploited/extended when novel metaphorical expressions are created and this is why newly and imaginative uses of the mapping are generally understood instantly. More specifically, successful new metaphorical meaning extensions are possible for words belonging to source domains of conceptual metaphors (cf. Alongs & Lönncker, 2004).

Of course, not all metaphors are universal, but a part of them are culturally determined; moreover, when “a conceptual metaphor is universal, its universality obtains at a generic level, while the same conceptual metaphor shows cultural variation at a specific level” (Kövecses, 2002: 248).

Complex metaphors are the result of the combination of primary metaphors, plus forms of commonplace knowledge, which can be related to specific cultures. Lakoff & Johnson (1999: 60-3) discuss the A PURPOSEFUL LIFE IS A JOURNEY complex metaphor, showing that it results from the combination of the cultural belief according to which people are supposed to have purposes in life and to pursue them, plus the PURPOSES ARE DESTINATIONS and ACTIONS ARE MOTIONS conceptual metaphors. Furthermore, we all know a simple fact, i.e. that “A long trip to a series of destinations is a journey”. Thus, complex metaphors are made up of different submetaphors; moreover, complex metaphors carry with them complex inferences.

While traditionally idioms were seen as “arbitrary”, according to the cognitive theory of metaphor most of them are motivated by metaphorical mapping, i.e. they are “metaphorical idioms” (see section 4 for examples).

3. Metaphor in NLP systems and in lexical resources

Due to the pervasiveness of metaphor in language, most NLP systems have to deal with it sooner or later. However, so far very few systems have been designed to handle specifically conventional metaphors. Even harder is, of course, finding systems able to deal with novel metaphorical expressions (but see KARMA features described in Narayanan, 1997; Feldman & Narayanan, 2004).

Existing systems are, in general, knowledge-based systems, because they rely on a representation of (at least a part of) the conventional links (mappings) between expressions in source domains and metaphorical expressions in corresponding target domains. These systems need to acquire a lot of knowledge, thus lexical resources providing rich data on metaphor can be very useful for them (but, of course, also for other kind of NLP systems).

For instance, Martin (1994) refers to the MIDAS project within which a set of computer programs had been developed to explicitly represent knowledge about conventional metaphors, apply this knowledge to interpret metaphorical expressions and also learn novel metaphorical expressions eventually found in texts. As Martin points out, the effectiveness of the MIDAS approach for language interpretation, generation and acquisition turned out to be obviously dependent on the quantity and quality of data of the knowledge-base available (and this was the case also for other systems mentioned in the paper). Thus, the paper describes the project ongoing to build MetaBank, a knowledge base containing empirically derived and theoretically motivated information on conventional metaphors in English. Also within the ATT-Meta project (Barnden & Lee, 2002), aimed at building a system able to perform the but there is a variation at specific levels of the metaphor, so whereas in English a linguistic expression of the metaphor can be “He is just blowing off steam”, the corresponding expression in Italian does not exist (i.e., is not conventional): “Lui sta liberando vapore” (but cf. “Il sangue gli ribolliva dalla rabbia” = “His blood was boiling with rage”).

2 Providing the possibility to say, e.g., “She is a warm person” or “They gave me a warm welcome”.
3 As Lakoff (1993: 206) puts is, if we say that “Our relationship has hit a dead-end street”, we implicate that “the relationship is stalled, that the lovers cannot keep going the way they’ve been going, that they must turn back, or abandon the relationship altogether”.
4 For instance, the already mentioned conceptual metaphor ANGER IS HEAT is instantiated both in English and in Italian,
reasoning needed for understanding metaphorical utterances (so far only related to people's beliefs), a database of metaphors is being built, containing data on conceptual metaphors and real discourse examples of them. The metaphor names under which examples are classified in the databank have been mostly invented by the ATT-Meta group, although there are some correspondences to the metaphor names used by other researchers (notably to the metaphor names used in the Master Metaphor List of Lakoff et al., 1991). Moreover, there are indications of hierarchical (but not clearly defined) relationships among metaphors.

The systems mentioned needed to build their own resources on metaphor because general lexical resources do not normally provide sufficient and consistent information on metaphorical expressions/metaphorical mechanisms. Martin (1994) also states that the semantic distinctions that these resources provide are generally too fine-grained and sometimes arbitrary. This was discussed also in various paper which dealt in particular with information encoded in wordnets, showing for instance that neither within the Princeton WordNet nor within EuroWordNet (EWN) the issue of how to treat metaphorical expressions was systematically dealt with. Alongo & Castelli (2002) emphasized the limits of EWN in this respect, by analysing data encoded within the Italian wordnet (further developed as ItalWordNet – IWN, Roventini et al., 2003). They showed that in IWN

a) information on metaphorial word senses is neither systematic nor consistent;
b) when information on metaphorical sense extensions is present, there is no indication of the connection between the “basic” and the “extended” senses;
c) data which could help identify novel metaphorical expressions in texts are not provided.

Thus, some proposals for the representation of metaphors in wordnets were put forward, while stressing that at the same time the necessity to add corpora as further sources for wordnets and to analyse them by adopting as a reference framework a well developed theory like the Cognitive theory of metaphor. Other works (Alongo & Castelli, 2003; Lönneker, 2003; Alongo & Lönneker, 2004) further deepened the issue and refined the proposals for a better encoding of data on metaphors in wordnets.7

The Hamburg Metaphor Database (HMD – Eilts & Lönneker, 2002 – available at http://www1.uni-hamburg.de/metaphern/index_en.html) is an online database of French and German metaphors which came into being in 2002. The goal of the researchers working at it is both showing the potential and overcoming the shortcomings of existing general lexical resources and specific metaphor databases. In the database, metaphors appearing in different domain-specific corpora of French and German, collected from mass media, are encoded. The metaphors are annotated with lexical and conceptual information, referring to the EWN database for lexical information and the Berkeley Master Metaphor List (Lakoff et al., 1991) for information on conceptual domains.

Apart from being obviously useful as a knowledge base for NLP systems which need to treat German and French metaphorical expressions, the HMD can be used to improve the quantity and quality of data within the German and French lexical networks in EWN. For instance, Lönneker (2004: 13) claims that when analysing data contained within the database or also when encoding data “it turns out very fast that the French and German lexical networks included in EWN have a rather low coverage even of some conventionalized metaphors”. Moreover, in EWN choices made with respect to metaphorical expressions often turn out to be arbitrary when compared to data in the HMD (cf. Lönneker, 2004). Thus, the information encoded in the HMD can be used both to fill gaps and to revise shortcomings and inconsistencies within the wordnets involved.

4. Structure and content of the IMD

The IMD is a resource designed to be used both as a specific repository of information on metaphors in Italian and to enrich existing general lexical resources, like e.g. IWN. The project was first inspired by the above mentioned research earlier carried out on the possibility of metaphor representation in lexical resources for NLP and in particular in wordnets (Alongo & Castelli, 2002, 2003). Therefore, the work started at the end of 2003, with the main aim of collecting (real) data on metaphors, in a principled and theoretically motivated way, given that no such data were available for Italian in existing resources. The theoretical framework used to structure and develop the database is the Cognitive theory of metaphor, briefly recalled above.

The first period of the project was devoted to design the database in such a way to foresee the encoding of (all) the information necessary to use it either to enrich computational lexical resources, or to directly use its data within NLP applications. Then, the encoding of the data begun, but during this second stage it has been sometimes necessary to modify choices made at the beginning, due to new insights gained from the analysis of data.

The database has been designed to provide information on the following aspects of metaphors:
- Cognitive metaphor
- Source domain
- Target domain
- (Hierarchical) link(s) with other metaphors
- Lexical expression(s) for the metaphor
- Multiword expressions for the metaphor (much work is being devoted at the study of idiomatic expressions linked to metaphors)
- Corresponding metaphor in English (the correspondence is looked for within the Berkeley Master Metaphor List)
- Metaphorical expressions eventually encoded in ItalWordNet
- Metaphorical multiword expressions eventually encoded in ItalWordNet
- Examples from a corpus of Italian.

---
5 http://www.cogsci.princeton.edu/~wn.
6 http://www.ilic.uva.nl/EuroWordNet.
7 The issue was also discussed in a Panel session at the 2nd Global WordNet Conference: see http://www.fi.muni.cz/gwc2004.
In order to identify a metaphor, different strategies are being followed: first of all, the conceptual metaphors distinguished for English within the Berkeley Master Metaphor List\(^8\) are taken into consideration. As we know, many metaphors are universally instantiated, so it can be the case that a conceptual metaphor identified for English is also found in Italian; however, even universal metaphors may show cultural variation at a specific level, thus it is necessary to look for expressions of the metaphor within a corpus of Italian. Moreover, the Berkeley database only lists some ad hoc created examples, and it is not necessarily representative of all the possible instantiations of a metaphor; thus, it is not sufficient to try to translate those examples into Italian. Therefore, for every metaphor listed in the Berkeley database, examples are looked for in the Parole Corpus of Italian (partly available online at \(\text{http://www.ilc.cnr.it/pisystem/demo/demo_dbt/demo_corpus/index.htm}\) – see Goggi et al., 2000), both to verify if it is present also in Italian and to find as many as possible instantiations of it, evaluating possible cultural variations. This comparison with the Berkeley database might also be useful to deal with metaphors in a multilingual perspective.

Other possibilities to identify metaphors in Italian are given by: directly taking in consideration texts taken from the corpus (given the pervasiveness of metaphors, each text analysed is a rich source of data); using results of theoretical works carried out specifically on Italian\(^9\); analysing data contained within lexical resources.

Once a cognitive metaphor is recognized as instantiated in Italian, an interesting issue to be tackled is that of identifying links with other metaphors. So far, we have only encoded opposition links, plus another generic link between metaphors, because no deep analysis of the relationships found has been carried out yet. For instance, a conceptual metaphor like NOTO È AVANTI (the known is ahead) has a reversal (IGNOTO È DIETRO – the unknown is at the back) which is instantiated in Italian, an interesting issue to be tackled is that of identifying links with other metaphors. Thus, for instance, a lot of multiwords/idiomatic expressions are not anomalous, but it is instead possible to identify regularities and motivations in their semantics.

An important aspect of the work being carried out is that we also check if lexical units and/or multiword expressions encoded in the IMD as metaphorical are properly encoded within IWN. As said above, much data on metaphorical expressions is missing or inconsistent in IWN/EWN (cf. Alonge & Castelli, 2002, 2003; Alonge & Lönneker, 2004), thus we are identifying data to fill in gaps or reorganize in a more consistent and theoretically motivated way information already existing in IWN. Furthermore, by connecting metaphorical lexical units and multiwords to a same metaphor, we aim at collecting as many as possible (or even all the) linguistic expressions found within a target domain of a conceptual metaphor. These data will also be useful to revise data encoded in IWN and the relationships among them. For instance, they might be used to detect which lexical units not encoded as conventional metaphorical expressions could be used as novel metaphors. In Alonge & Lönneker (2004) the case of gestazione and gravidanza (both meaning pregnancy, gestation) was discussed. They are found together in a literal meaning synset in IWN, and are related to nascere (to be born), which has both a literal and a metaphorical sense, by means of a (non-factive) cause relation. Gestazione has, actually, also a metaphorical sense encoded in the wordnet, and in fact examples of it were found in the corpus analysed (“Ignoravo che il decreto fosse in gestazione” = I ignored that the decree was in gestation (in progress)). However, the metaphorical synset does not contain the synonym gravidanza: as it is the case for other words, this could be a shortcoming/inconsistency of the database. However, by analysing a corpus of Italian we did not find any metaphorical usage of the word. Thus, gravidanza is not conventionally used in a metaphorical sense, although by searching in the Internet the following example was found:

- Una piacevole, impegnativa e interessante gravidanza, iniziata due anni fa, ha avuto come esito questo libro

(A pleasant, engaging and interesting pregnancy, started two years ago, has had this book as a result).

This is in line with the Cognitive theory of metaphor, according to which novel metaphorical sense are generally created by extending meanings of words found in the same source domain of a metaphor. Thus, gravidanza, although not conventionally metaphorical, but found in the same source domain of gestazione, might be sometimes used in a metaphorical sense without determining understanding problems. In the IMD gravidanza is not encoded as metaphorical: by using this information, in IWN it will be possible to correctly revise different conceptual metaphors, it is shown that idiomatic expressions are not anomalous, but it is instead possible to identify regularities and motivations in their semantics.

\(^8\) These can also be found online at the Conceptual Metaphor Homepage, \(\text{http://cogsci.berkeley.edu/lakoff/}\), where indices by metaphor names, source domains, and target domains are available.

\(^9\) In particular, we are using a very rich study provided by Casadei (1996) on idiomatic expressions in Italian and their correlations with conceptual metaphors.
data and also encode information on the possible metaphorical use of the word, by means of the mechanisms proposed by Alonge & Castelli (2002) and Alonge & Lönneker (2004).

5. Conclusion and future work

In this paper we have described the main features of the IMD, a resource on conceptual/linguistic metaphors in Italian, being developed at the University of Perugia and which will soon be made available for research by academic institutions.

So far 97 conceptual metaphors have been richly encoded in the database, most of which have a correspondence within the Berkeley database. However, the links encoded among metaphors within IMD are not systematically indicated within the Berkeley database, which misses also systematic data on lexical units and multiword expressions connected with a conceptual metaphor.

At the moment, the metaphors encoded are exemplified by 473 multiword/idiomatic expressions and 260 lexical units, for which a correspondence has been looked for within IWN. Further work needs to be devoted to the insertion of corpus examples, which, so far, have been only occasionally added.

In the future we plan to further enrich the database: in a first stage we are going to add more lexical/multiword expressions for already listed metaphors, and examples from the corpus. A second stage will be devoted to also add, and richly encode, further metaphors.

6. References

Alonge, A. and Castelli, M. (2002). Which Way Should We Go? Metaphoric Expressions in Lexical Resources. In Proceedings of the 3rd International Conference on Language Resources and Evaluation. Las Palmas, Canary Islands, Spain, pp. 1948–1953.

Alonge, A. and Castelli, M. (2003). Encoding Information on Metaphoric Expressions in WordNet-like Resources. In Proceedings of the Workshop on The Lexicon and Figurative Language - 41st Annual Meeting of the Association for Computational Linguistics, Sapporo, Japan.

Alonge, A. and Lönneker, B. (2004). Metaphors in Wordnets: from Theory to Practice. In Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC), 24-30 Lisbon, Portugal, pp. 165-8.

Barnden, J.A. and Lee, M.G. (2002). An Artificial Intelligence Approach to Metaphor Understanding. In Tomasz Komendzinski (ed.), Metaphor: A Multidisciplinary Approach, special issue, Theoria et Historia Scientiarum, 6 (1), pp.399-412.

Casadei, F. (1996). Metafore ed espressioni idiomatiche. Uno studio semantico sull’italiano. Roma, Bulzoni.

Eilts, C. and Lönneker, B. (2002). The Hamburg Metaphor Database. In Metaphorik.de 03/2002, pp. 100-110.

Feldman, J. and Narayanan, S. (2004). Embodied Meaning in a Neural Theory of Language. Brain and Language, 89, pp. 385-392.

Goggi, S., L. Biagini, E. Picchi, R. Bindi, S. Rossi, R. Marinelli, (2000). Italian Corpus Documentation. LE-PAROLE, WP2.11.

Grady, J. (1997). Foundations of Meaning: Primary Metaphors and Primary Scenes. Ph.D. Dissertation, Berkeley, University of California.

Johnson, C. (1997). Metaphor vs. Conflation in the Acquisition of Polysemy: The Case of SEE. In M. K. Hiraga, C. Sinha, and S. Wilcox (eds.), Cultural, Typological and Psychological Issues in Cognitive Linguistics. Current Issues in Linguistic Theory 152. Amsterdam: John Benjamins.

Kövecses, Z. (2002). Metaphor. A Practical Introduction. New York: OUP.

Lakoff, G. (1993). The Contemporary Theory of Metaphor. In Ortony, Andrew (ed.) Metaphor and Thought, 2nd. ed., Cambridge University Press, pp. 202-251.

Lakoff, G., Espenson, J. and Schwartz, A. (1991). Master metaphor list. Draft 2nd Edition. Cognitive Linguistics Group, University of California at Berkeley, CA.

Lakoff, G. and Johnson, M. (1980). Metaphors We Live by. Chicago: University of Chicago Press.

Lakoff, G. and Johnson, M. (1999). Philosophy in the Flesh. The Embodied Mind and its Challenge to Western Thought. New York, Basic Books.

Lönneker, B. (2003). Is there a Way to Represent Metaphors in WordNets? Insights from the Hamburg Metaphor Database. In Proceedings of the Workshop on The Lexicon and Figurative Language, 41st Annual Meeting of the ACL. Sapporo, Japan, pp. 18–26.

Lönneker, B. (2004). Lexical Databases as Resources for Linguistic Creativity: Focus on Metaphor. In Veale, T., Cardoso, A., Camara Pereira, F. and Gervás, P. (eds.), Language Resources for Linguistic Creativity, LREC 2004 Satellite Workshop, pp. 9-16.

Martin, J.H. (1994). MetaBank: a Knowledge-Base of Metaphoric Language Conventions. Computational Intelligence, 10 (2), pp. 134-149.

Narayanan, S. (1997). Knowledge-based Action Representations for Metaphor and Aspect (KARMA). Ph.D. Dissertation, Computer Science Division, University of California, Berkeley.

Roventini A., Alonge A., Bertagna F., Calzolari N., Cancila J., Girardi C., Magnini B., Marinelli R., Speranza
M., Zampolli A. (2003). ItalWordNet: Building a Large Semantic Database for the Automatic Treatment of Italian, in A. Zampolli, N. Calzolari, L. Cignoni (eds.), *Computational Linguistics in Pisa*, Special Issue, Vol. XVI-XVII, Pisa, Giardini, pp. 745–792.