Disambiguation of Basic Action Types through Nouns' Telic Qualia

Irene Russo, Francesca Frontini, Irene De Felice,
Fahad Khan, Monica Monachini
ILC CNR / Via G. Moruzzi 1, 56124 Pisa
{irene.russo,francesca.frontini,irene.defelice,
fahad.khan,monica.monachini}@ilc.cnr.it

Abstract

Knowledge about semantic associations between words is effective to disambiguate word senses. The aim of this paper is to investigate the role and the relevance of telic information from SIMPLE in the disambiguation of basic action types of Italian HOLD verbs (*prendere*, 'to take', *raccogliere*, 'to pick up', *pigliare* 'to grab' etc.). We propose an experiment to compare the results obtained with telic information from SIMPLE with basic co-occurrence information extracted from corpora (most salient verbs modifying nouns) classified in terms of general semantic classes to avoid data sparseness.

1 Introduction

Word senses emerge in lexicographic practice as the result of splitting strategies depending on context of use, syntagmatic patterns and perceived semantic similarity. Lexicographers share working assumptions (e.g. the concrete sense is encoded before the abstract sense of a lemma) on the way to structure glosses. The induction of word senses from corpus co-occurrences, as in the Corpus Pattern Analysis effort (Hanks 2008), has an impact on the definition of how many different senses are available and, with the focus on general semantic classes of nouns involved for example as objects in verbal contexts, the path toward sense induction is made fully empirical.

Since word senses are not metaphysical objects but depend on dedicated tasks that require them (Kilgarriff 1997), other operative principles are possible. In this paper we present a manually annotated dataset relative to basic Italian action verbs that have been partitioned in basic action types when the action described in the sentence was analysed in terms of body movements involved. This split among senses can't be unequivocally aligned with lexical resources such as WordNet (Moneglia et al. 2012) and, even if the induction from corpora examples implies that the syntagmatic structure is important, the guiding motivation in segmenting the meaning concerns salient differences in the action performed by the agent for the sake of basic action modelling in robotics.

In this dataset a central role is assigned to nouns denoting concrete objects and as a consequence the task of basic action type classification focuses on nouns and the information attached to them that could help in disambiguation.

In word sense disambiguation tasks different sets of features have been tested in order to understand which are the most relevant for classifying senses. Among these features, there are PoS and syntactic information, collocations (or selectional preferences), thematic roles, semantic associations between words in terms of taxonomic relations (e.g. *chair*, *furniture*), events (e.g. *chair*, *sitting*), topic (e.g. *bat*, *baseball*), head argument relations (e.g. *dog*, *bite*). (Agirre and Martinez 2001) reviewed these features, discovering that collocations and semantic associations are the most useful (and manually annotated corpora are the best source to acquire them).

The aim of this paper is to investigate the role and the relevance of telic information from SIMPLE (Ruimy et al. 2003) in the disambiguation of basic action types of Italian HOLD verbs (*prendere*, 'to take', *raccogliere*, 'to pick up', *pigliare* 'to grab' etc.). We propose an experiment (see 5) to compare the results obtained with telic information from SIMPLE with basic co-occurrence information extracted from corpora (most salient verbs modifying nouns) classified in terms of general semantic classes to avoid data sparseness.

2 ImagAct Basic Action Types: Bottom-up Derivation of Verbs Senses

Action verbs are among the most informative elements in a sentence: the concepts they codify have a great relevance in human life and they are the most frequent elements in speech (Moneglia and Panunzi, 2007). In our everyday experience, the kind of actions we can carry out is almost endless but, given that every human language tends towards economy of expression, the number of action verbs we use is always somehow restricted. So we adopt the same verbs to denote different types of events: for example, the verb "to take" in (1) *John takes a present from a stranger* means "to receive, to accept"; but in (2) *John takes Mary the book* it means "to bring"; in (3) *John takes the pot by the handle* it simply means "to grasp"; finally, in (4) *John takes Mary to the station* it means "to conduct, to accompany". Furthermore, every language manifests a different behaviour in segmenting human experience into its proper action verbal lexicon. For this reason, the examples just cited can't be translated with a single Italian verb: (1a) *John prende un regalo da uno straniero*; (2a) *John porta il libro a Maria*; (3a) *John prende la tazza dal manico*; (4a) *John porta Maria alla stazione*. But we expect that, in a given language, similar events will be referred to by using the same verb: so "to take" will apply also to *John takes the children to school/his wife to the cinema*, similarly to (4); we also expect this tendency to be found in other languages, as is the case for "portare" in *John porta i bambini a scuola/sua moglie al cinema*, similar to (4a). In the ImagAct framework these coherent sets of similar events are referred to as action types. Verbs which extensionally denote more than one action types (as "to take") are named general verbs.

Since written corpora tend to abound in abstract verbs or verbs used in their abstract senses, the best way to study action verbs' actual variation is in spontaneous speech, i.e. in transcribed spoken corpora. The ImagAct project focuses on high frequency action verbs (approximately 600 lexical entry) of Italian and English, which represent the basic verbal lexicon of the two languages. All occurrences of these verb were retrieved, respectively from a collection of Italian spoken corpora (C-ORAL-ROM; LABLITA; LIP; CLIPS), and from the BNC-Spoken; linguistic contexts of each occurrence were then standardized and reduced to simple sentences as those reported above (1-4).

Once the ImagAct corpus was created, as a first step, annotators made a distinction between the metaphorical and phraseological usages (e.g. *John takes Mary to be honest*) from proper occurrences of action verbs (e.g. *John takes the glass*); then, they grouped the occurrences into action types, keeping granularity to its minimal level, so that each type contained a number of instances referring to similar events (*John takes the glass/the umbrella/the pen* etc.). This procedure was accomplished through a web based annotation interface and was standardized in the specifications of the ImagAct project. Finally, one best example was chosen (or more than one, if the verb had more than one possible syntactic structure) from all standardized sentences of each type, and it was then associated to a video to exemplify the action type.

To obtain a parallel corpus, all English standardized instances assigned to a type have been translated into Italian, and vice versa; the possibility of translating all instances of a type into another language, using only one verb, assures the coherence of that type. In the very last phase, a mapping between English and Italian action types has been conducted onto the same set of scenes. The validation of basic action types is going on also for Chinese and in the future other languages will be involved in this procedure. Crosslinguistic comparison between languages highlights coarse-grained distinctions between word senses (Resnik and Yarowsky 1998); as a consequence we expect that the extension to more languages will make the basic action types more general and less dependent on a specific language.

The result of the procedure described above is a set of short videos, each one corresponding to an action type and showing simple actions (e.g. a man taking a glass on a table), by which a user can access the English/Italian best examples chosen for that type (*John takes the glass/John prende il bicchiere*) and all the standardized sentences extracted from corpora that have been assigned to that type; these videos show the actual use of the verb when referring to a specific type of action. Also, a user can access this data by lemma: for example, searching for the verb "to take", he will be presented with a number of scenes, showing the different action types associated to that verb, with their related information. Scenes, and their associated best examples, represent the variation of all action verbs considered and constitute the ImagAct ontology of action. This ontology is not only inherently inter-

linguistic, having been derived through an inductive process from corpora of different languages, but also takes into account the intra-linguistic and inter-linguistic variation that characterizes action verbs in human languages.

3 GL Co-Composition and ImagAct General Verbs

Pustejovsky (1995) defines co-composition as a semantic property of a structure in which both a predicate and its argument(s) contribute functionally to the meaning of an expression, so that the semantic contribution of the argument(s) of the predicate is greater than can be accounted for on a strictly compositional analysis of meaning. We can then view certain verbs as being lexically underspecified in the sense that the arguments of these verbs play a significant role in ascertaining the full meaning of the verb in context. The classic example of co-compositionality as given in Pustejovsky (1995) involves the verb "bake" which can be understood in at least two distinct senses, a "change of state" sense as in sentence (5) and a "creation" sense as in sentence (6):

(5) John baked the potato.
(6) John baked the cake.

This can be understood as an example of logical polysemy since, although "bake" has a slightly different meaning in each of the two sentences (5) and (6), these meanings are somehow closely related. It's not hard to find other examples in which co-compositionality is clearly evident and where the meaning of a verbal predicate in context, including the type of action to which it might refer, is heavily dependent on the type and meaning of its arguments. So that for example the following two sentences refer to two different action types for the Italian verb *prendere* in ImagAct:

(7) *Marco prende la mela*. ('Marco takes the apple')
(8) *Marco prende la mela dall'albero*. ('Marco picks the apple from the tree').

One could argue that those action verbs which best fit the definition of lexical underspecification as given above should also be regarded as "general" verbs in the ImagAct sense. Each general verb is associated with a finite set of action types, which are themselves determined by the different kinds of objects to which the actions referred to by the verb might apply. If an action verb is underspecified, it can be said to lack one determinate meaning which might provide a clear prototypical example of the kinds of actions to which it refers, so that a verb like "to open" or "to take" is "vague" enough to be associated with a number of distinctive action types in its primary non-metaphorical usages, whereas a verb like "to knife" or even "to eat" can plausibly be associated with only one type of action: this is at least the viewpoint taken up the ImagAct project (www.imagact.it). Thus, ImagAct could be viewed as an important lexical resource for the analysis of the phenomena of co-composition at least to the extent that it pertains to the class of action verbs.

4 Enriching HOLD Verbs' Sentences with Semantic Information from SIMPLE

A disambiguation task performed on manually annotated data involving action types has a practical application, considering that these data will be analysed in the on-going ModelAct project for human-robot interaction and modeling of actions. However, the results have also theoretical implications because the way the senses have been individuated is peculiar and the kind of meanings classified (verbs' senses referring to concrete actions) can change the expectations about the most relevant/useful knowledge source for disambiguation. In this paper we mainly use semantic associations knowledge from SIMPLE to disambiguate between basic action types.

The Italian component of the ImagAct dataset contains at the moment 744 verbs and 1358 basic action types, for a total of 26233 standardized sentences. The intra-linguistic mapping between basic action types to discover local equivalence between verbs is in progress. In this paper we focus on a semantically coherent verbs' class, that of Levin's HOLD verbs (Levin 1993) (*to clasp, to clutch, to grasp, to grip, to handle, to hold, to wield*), corresponding to Italian verbs *acchiappare, afferrare agguantare, pigliare, prendere, raccattare, raccogliere, stringere, tenere.* Looking at basic action types of these verbs, we find several equivalence *(Marco piglia lo yogurt dal frigorifero/ Marco prende il prodotto dalla busta*) that will be grouped in the disambiguation experiment (see 5).

We extract from SIMPLE the telic information about the objects of the HOLD verbs.

We decided to use SIMPLE because of great amount of structured encyclopedic knowledge it contains. SIMPLE is largely based on Pustejovsky's Generative Lexicon (GL) theory. GL theory posits that the meaning of each word in a lexicon can be structured into components, one of which, the qualia structure, consists of a bundle of four orthogonal dimensions.

These dimensions allow for the encoding of four separate representative aspects of the meaning of a word or phrase: the formal, namely that which allows the identification of an entity, i.e., what it is; the constitutive, what an entity is made of; the telic, that which specifies the function of an entity; and finally the agentive, that which specifies the origin of an entity. These qualia structures play an important role within GL in explaining for example, the phenomena of polysemy in natural languages. SIMPLE itself is actually based on the notion of an extended qualia structure, which as the name suggests is an extension of the qualia structure notion found in GL. Thus, there is a hierarchy of constitutive, telic, and agentive relations that can hold between semantic units. SIMPLE contains a language independent ontology of 153 semantic types as well as 60k so called "semantic units" or USems, representing the meanings of lexical entries in the lexicon. SIMPLE also contains 66 relations organized in a hierarchy of types and subtypes all subsumed by one of the four main qualia roles:

- FORMAL (is-a)
- CONSTITUTIVE, such as ACTIVITY produced-by
- TELIC, such as INSTRUMENTAL used-for
- AGENTIVE, such as ARTIFACTUAL caused-by

4.1 Manual annotation of affording properties

Since HOLD verbs selected are the most generic verbs involving actions done with hands, a manual annotation has been done on each sentence in terms of affording properties of the objects (Gibson 1979).

As additional information we annotated the properties of the objects denoted by lemmas that afford grasping. These properties are defined by the type of grasping the object afford. We created these categories adopting a bottom-up approach, by looking at all the possible objects of primary verbs and identifying a minimum set of common features among them.

One-Hand_Grasp: this is a property of objects that can be grasped using only one hand. The size of two of the object's dimensions (length, width or thickness) must not exceed the maximum span of a hand with at least two fingers bent in order to grasp and hold something. E.g.: "Johs takes the lighter". The agent's control over the grasped object is maximum.

Two-Hands_Grasp: this property is still related to the object size and qualifies objects that cannot be grasped without necessarily using two hands. Note this kind of grasp is not specifically directed to any of the object's parts. E.g.: "John takes the board". Also in this case, the agent's control over the grasped entity is very high, also with animates (when they can be taken and hold with two hands, as in "The nurse takes the baby from the incubator").

Grasp_by_part: this property is proper of big objects (i.e., whose size exceed the maximum span of a hand) that, even so, can be perfectly controlled by agents using only one hand thanks to a handle. Handle refers here to any part of an object specifically designed to afford grasping (like a handle of a handbag). This property is also shown by objects with dimensions bigger than a hand size, especially all animate entities, that have no designed handles, but that still can be grasped and hold simply using one hand: the grasp will be directed to one of their parts, the one (usually hand, arm) that better allows grasping for its suitability in size with hands (but note that in these cases agent's control over the grasped entity is much less strong). These parts are often explicitly mentioned for their relevance for action (especially if there are many possible graspable parts in the same entity, as in "John takes Mary by her hand/her leg/her arm"), for they are not predetermined, as designed handles are.

Grasp_with_instrument_container: this is the main property of entities (mainly substance and mass entities) which humans cannot directly control without using some other object, because of their fluid consistency and because of the absence of a solid, tangible, definite shape contour. For example, water and other liquids cannot be grasped without a container, as a bottle or a glass. Because it is impossible for humans to grasp these entities without a recipient, explicit reference to the container is often omitted (as in "John takes the water for the dog from the faucet": it is implicitly understood that he uses a

bowl), and in some cases is even lexicalized, as demonstrated by the fact that some objects can accept a quantified form, as in "John takes two beers out of/from the fridge" (= bottles of beer). In this example, the grasping event properly involves the solid container, but is semantically referred to the content. This kind of polysemy (container/content), which originates from metonymic processes, is quite regular and widespread in languages: this can be easily understood considering that, for humans, contents are usually much more salient than containers.

Additional information: sometimes, objects shape, dimensions and constituency do not suffice to predict how humans actually grasp them. For this reason, we annotated some objects with two affording properties. This mainly concerns objects that can be grasped with one hand, but that usually are grasped with an instrument (that in turn can be grasped with one or two hands). For example, *zucchini*, *potatoes*, *meat* and other foods (as in "John takes the zucchini/the meatball from the tray"), can be grasped directly by hands, but usually we prefer to use a fork (grasped with one hand). So, for *zucchini*, when intended as [food], we annotated both one_hand_grasp and one_hand_instrumental_grasp. Another case in which we annotated two affording properties is when an object is one_hand_graspable, but it has a part specifically designed for grasping (as for scissors or pacifiers). In this case, we annotated both one_hand_graspabable and grasp_by_part.

5 Disambiguation of HOLD Verbs Basic Action Types: a First Experiment

Our starting dataset comprises 1419 sentences and 29 basic action types. Some sentences were doubled because the telic qualia in SIMPLE for several nouns has more than one entry. At the end we have 1573 instances to classify. We performed a ten-fold cross validation experiment with the implementation provided in WEKA (Hall el al. 2009) of Support Vector Machine algorithm (called SMO) since results from the literature WSD on benchmark data show that support vector machines (SVMs) yield models with one of the highest accuracies.

The features for this experiment are:

- manually annotated information about the semantic class of nouns in WordNet 3.0 (SCN in Table 1):
    - *libro* ('book') → artifact
    - *caramelle* ('candies') → cibo
- annotation on the affording properties of objects as described in 4.1 (AffP in Table 1).
- values encoded for Telic qualia in SIMPLE, manually disambiguated for each noun and reported as Boolean values for each of the 23 verbs' abstract semantic classes in SIMPLE (as Cause_Constitutive_Change in the following example) (SIMPLE in Table 1):

  Matteo prende il coltello. 'Matteo takes the knife'
  knife **UsedFor** tagliare ('to cut')

  tagliare is Cause_Constitutive_Change
- SIMPLE semantic classes of most salient verbs that precede the target noun in itTenTen, a web corpus of 3.1 billion tokens, accessible through APIs provided by sketchengine.co.uk . These data have been extracted as word sketches (Kilgarriff et al. 2004) and as a consequence report on selectional preferences that are among the most useful features in WSD (see Introduction) (itTenTen in Table 1). We found out that this pattern extracts content similar to telic qualia and for this reason we compare telic information and word sketches in this experiment.

We also perform disambiguation experiment on a version of the dataset with grouped action types (i.e. 14) composed by 1577 sentences because we found equivalence between several types. Baseline assigns each sentence to the most common action type (75.3% for AllBT, the dataset with all the basic action types, and 84% for GroupedBT, the dataset with grouped action types).

The results are reported in Table 1.

	AllF	SCN	AffP	SIMPLE	itTenTen
AllBT	81.6%	77%	76%	77.4%	80.5%
GroupedBT	89.7%	86.9%	85.7%	87.6%	90.2%

Table 1: Accuracy for basic action types disambiguation with different set of features

The best result is obtained on the grouped basic action types dataset, with 0.88 as preci-

sion and 0.90 as recall. For this dataset information extracted from SIMPLE have a small negative impact on the accuracy while for the dataset with all the action types it contributes to improve the result. Affording properties are not very relevant for disambiguation: even if the affordances of objects are known from psychological studies as a relevant feature in action learning, the annotation proposed is probably not the best way to represent this knowledge.

6 Conclusions and Future Work

Knowledge about semantic associations between words is effective to disambiguate word senses. Distributional models of word meanings represent this information providing a vector-based representation of most frequent words in context. We extracted this information from SIMPLE, a rich lexical resource that provide essential information about objects' typical uses in the telic qualia. The three most salient verbs that have as object the target nouns in ImagAct sentences have been extracted from a large web corpus. To avoid data sparseness SIMPLE complex ontology that label verbs with coarse-grained semantic classes have been used. The results show that qualia information is useful for disambiguation but enriching it with salient data from corpus improves the accuracy.

As future work we want to enrich the ImagAct dataset with information from other qualia in SIMPLE (i.e. formal, constitutive and agentive) and from other resources, such as dictionary's glosses, ontologies for actions, distributional data from different corpora with the aim to find the best set of features for the disambiguation of basic action types. As a collateral project, we plan to find additional salient values for nouns' qualia structure through patterns in corpora.

References

Agirre, E. Martinez, D. (2001), Knowledge sources for word sense disambiguation. In Proceedings of International Conference on Text, Speech and Dialogue (TSD'2001) Selezna Ruda, Czech Republic.

Gibson, J. J. (1979). The Ecological Approach to Visual Perception. Boston: Houghton Mifflin.

Hall, M. Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.

Hanks, Patrick. (2008). "Mapping meaning onto use: a Pattern Dictionary of English Verbs". AACL 2008, Utah.

Levin, B. (1993), "English Verb Classes and Alternations: A Preliminary Investigation." In: The University of Chicago Press.

Kilgarriff, A.; Rychly, P.; Smrz, P.; Tugwell, D. (2004). 'The Sketch Engine'. Proceedings Euralex. Lorient.

Kilgarriff, A. (1997), "I don't believe in word senses" Computers and the Humanities 31: 91-113.

Moneglia, M., Alessandro Panunzi, Gloria Gagliardi, Monica Monachini, Irene Russo (2012), Mapping a corpus-induced ontology of action verbs on ItalWordNet. Global Wordnet Conference 2012.

Pustejovsky, J. (1995), The Generative Lexicon. MIT Press, Cambridge, MA.

Resnik, P., Yarowsky, D. (1998) Distinguishing Systems and Distinguishing sense: new evaluation methods fot WSD. Natural Language Engineering.

Ruimy, N, M. Monachini, E. Gola, N. Calzolari, M.C. Del Fiorentino, M. Ulivieri, and S. Rossi (2003), A computational semantic lexicon of italian: SIMPLE. Linguistica Computazionale XVIII-XIX, Pisa, pages 821–64.