

# A Resource and Tool for Super-sense Tagging of Italian Texts

G. Attardi<sup>1</sup>, S. Dei Rossi<sup>1</sup>, G. Di Pietro<sup>2</sup>, A. Lenci<sup>2,3</sup>, S. Montemagni<sup>3</sup>, M. Simi<sup>1</sup>

<sup>1</sup>Dipartimento di Informatica, Università di Pisa  
Largo B. Pontecorvo 3, I-56127 Pisa, Italy

<sup>2</sup>Dipartimento di Linguistica, Università di Pisa  
via S. Maria 36, I-56127 Pisa, Italy

<sup>3</sup>ILC-CNR, Pisa  
via G. Moruzzi 1, I-56124 Pisa, Italy

E-mail: attardi@di.unipi.it, deirossi@di.unipi.it, julia648@gmail.com, alessandro.lenci@ling.unipi.it, simonetta.montemagni@ilc.cnr.it, simi@di.unipi.it

## Abstract

A SuperSense Tagger is a tool for the automatic analysis of texts that associates to each noun, verb, adjective and adverb a semantic category within a general taxonomy. The developed tagger, based on a statistical model (Maximum Entropy), required the creation of an Italian annotated corpus, to be used as a training set, and the improvement of various existing tools. The obtained results significantly improved the current state-of-the art for this particular task.

## 1 Introduction

Super-senses were introduced by Ciaramita and Johnson (2003) for classifying common nouns according to 26 semantic categories, corresponding to the nouns subset of the lexicographer classes of WordNet (Fellbaum, 1998). These lexical-semantic broad categories are used in WordNet to group semantically close synsets into a coarse-grained ontology in order to facilitate the lexicographers' work of updating and managing the database.

In Ciaramita and Altun (2006) the classification was extended to include verbs (classification wrt 41 broad categories). In the same paper, Super-Sense Tagging (SST) was defined as a task half-way between Named-Entity Recognition (NER) and Word Sense Disambiguation (WSD). It can be considered as an extension of NER since it classifies a wider class of nouns (and verbs) in a larger set of general semantic categories. On the other hand, it is an easier and more practical task with respect to WSD, since WordNet defines very specific senses (strictly related to each synset) and it is quite difficult to infer the correct one from the context.

SST can therefore be of practical value in a number of NLP tasks involving world knowledge such as semantic information retrieval, question answering and information extraction. Moreover the small size of the ontology allows considering the problem as a sequence labeling task and exploiting machine learning techniques.

Ciaramita and Altun (2006) also describe a tool for super-sense tagging based on a discriminative HMM, trained with an average perceptron algorithm (Collins, 2002). Their tagger was trained on portions of the SemCor corpus (Miller et al., 1993) and still represents the state of

the art for English, with an average F-Score<sup>1</sup> of 77.18 (Table 1), and higher F-scores (up to 89) on the most frequent categories such as "person".

In this paper we describe how we created a training corpus for Italian annotated with super-senses and present the design of a tagger capable of achieving accuracies for Italian comparable to those for English.

## 2 SST for Italian

Picca et al. (2008) developed an SST for Italian using the tagger by Ciaramita and Altun, trained on MultiSemCor (Bentivogli et al., 2004), a version of SemCor, obtained by word-by-word translation into several languages including Italian.

Each translated word is aligned to its English counterpart and inherits its connection to a WordNet synset. The semantic information related to a term can then be obtained from WordNet, including its super-sense.

The accuracy achieved with these settings was much lower than for English and is reported in Table 1.

	Precision	Recall	F1
English	76,65	77,71	77,18
Italian Picca et al.	62,26	63,57	62,90

Table 1: Performance on MultiSemCor: Italian – English

We also experimented using the same settings, trying to investigate the reasons for the large difference in accuracy between English and Italian. Analyzing the quality of the training resource revealed that it suffered from various problems that can be summarized as follows:

<sup>1</sup> F-score is a standard accuracy measure in classification, defined as the weighted average of *precision* and *recall*.

- *Smaller size*: only 64% of the English corpus was translated to Italian and aligned.
- *Incomplete super-sense alignment*: several terms that had an associated super-sense in English did not have one in the Italian translation. This was due in part to the translation process that ignored foreign words (e.g. “Grand”, “Jury”, “Fulton” in Listing 1), in part because English multiwords had no Italian equivalent (e.g. “primary\_election” → “elezione primaria” in Listing 1).
- *Lack of representativeness*: the corpus consists of English texts of the ’60s translated to Italian and so it contains many obsolete or unusual words. Moreover sentences have an artificial or uncommon syntactic structure since they are the result of a literal translation.
- *Coarseness of Part-of-Speech*: PoS tags do not include morphological information.

Venerdì	S	B-noun.time
il	RD	0
Grand	SP	0
Jury	SP	0
di	SP	0
Fulton	SP	0
ha	VA	0
detto	V	B-verb.communication
che	CS	0
un'	RI	0
indagine	S	B-noun.cognition
sulla	EA	0
recente	A	B-adj.all
elezione	S	0
primaria	A	0
(...)		

Listing 1: First sentence of Italian MultiSemCor

To improve the quality of the annotation, we used a modified version of the Hunpos tagger (Halácsy, Kornai & Oravecz, 2007; Attardi et al., 2009) to re-tag automatically (with an accuracy of about 97%) the text with part-of-speech. In this process we added morphological information that represents an important feature for Italian.

Unlike the previous experiment in (Picca et al., 2008), we applied the SST developed by Ciaramita and Altun to classify according to 45 super-senses, corresponding to the whole set of the WordNet lexicographer classes, including adjectives and adverbs, in addition to nouns and verbs (the complete list of super-senses and a short description is shown in Table 2).

Id	SuperSense	Description
00	adj.all	<i>all adjective clusters</i>
01	adj.pert	<i>relational adjectives (pertainyms)</i>
02	adv.all	<i>all adverb</i>
03	noun.Tops	<i>unique beginner for nouns</i>
04	noun.act	<i>nouns denoting acts or actions</i>
05	noun.animal	<i>nouns denoting animals</i>
06	noun.artifact	<i>nouns denoting man-made objects</i>
07	noun.attribute	<i>nouns denoting attributes of people and objects</i>

08	noun.body	<i>nouns denoting body parts</i>
09	noun.cognition	<i>nouns denoting cognitive processes and contents</i>
10	noun.communication	<i>nouns denoting communicative processes and contents</i>
11	noun.event	<i>nouns denoting natural events</i>
12	noun.feeling	<i>nouns denoting feelings and emotions</i>
13	noun.food	<i>nouns denoting foods and drinks</i>
14	noun.group	<i>nouns denoting groupings of people or objects</i>
15	noun.location	<i>nouns denoting spatial position</i>
16	noun.motive	<i>nouns denoting goals</i>
17	noun.object	<i>nouns denoting natural objects (not man-made)</i>
18	noun.person	<i>nouns denoting people</i>
19	noun.phenomenon	<i>nouns denoting natural phenomena</i>
20	noun.plant	<i>nouns denoting plants</i>
21	noun.possession	<i>nouns denoting possession and transfer of possession</i>
22	noun.process	<i>nouns denoting natural processes</i>
23	noun.quantity	<i>nouns denoting quantities and units of measure</i>
24	noun.relation	<i>nouns denoting relations between people or things or ideas</i>
25	noun.shape	<i>nouns denoting two and three dimensional shapes</i>
26	noun.state	<i>nouns denoting stable states of affairs</i>
27	noun.substance	<i>nouns denoting substances</i>
28	noun.time	<i>nouns denoting time and temporal relations</i>
29	verb.body	<i>verbs of grooming, dressing and bodily care</i>
30	verb.change	<i>verbs of size, temperature change, intensifying, etc.</i>
31	verb.cognition	<i>verbs of thinking, judging, analyzing, doubting</i>
32	verb.communication	<i>verbs of telling, asking, ordering, singing</i>
33	verb.competition	<i>verbs of fighting, athletic activities</i>
34	verb.consumption	<i>verbs of eating and drinking</i>
35	verb.contact	<i>verbs of touching, hitting, tying, digging</i>
36	verb.creation	<i>verbs of sewing, baking, painting, performing</i>
37	verb.emotion	<i>verbs of feeling</i>
38	verb.motion	<i>verbs of walking, flying, swimming</i>
39	verb.perception	<i>verbs of seeing, hearing, feeling</i>
40	verb.possession	<i>verbs of buying, selling, owning</i>
41	verb.social	<i>verbs of political and social activities and events</i>
42	verb.stative	<i>verbs of being, having, spatial relations</i>
43	verb.weather	<i>verbs of raining, snowing, thawing, thundering</i>
44	adj.ppl	<i>participial adjectives</i>

Table 2: WordNet lexicographer classes (super-senses)

The best result we achieved with this setting was an average F-score of 64.95 (Table 3), which represents a small improvement on the baseline result but it is still not comparable with the performance on the English language.

	Precision	Recall	F1
Italian <small>Picca et al.</small>	62,26	63,57	62,90
Italian <small>Our</small>	64,95	64,95	64,95

Table 3: Performance on Italian MultiSemCor

The quality of the corpus could hardly be improved further without investing significant time to complete it manually; furthermore some problems and limitations described above could not be overcome.

Our goal was to use the SST tagger as a practical tool within a project, called SemaWiki (Attardi et al., Semawiki), that aims at developing NLP technologies for analyzing texts in Italian. The suite of linguistic tools developed, called TanL (Text Analytics and Natural Language), has been used to analyze the Italian Wikipedia for uses such as question answering, information extraction and semantic retrieval. So we were interested in the performance of the tagger on a real Italian corpus, in its integration in the linguistic pipeline of the project, the TANL pipeline (Attardi, Dei Rossi & Simi, 2010), as well as efficiency.

To achieve this purpose, we proceeded along two different directions:

- producing a better and more representative resource for training SST taggers for Italian;
- developing a new SST tagger, based on the Maximum Entropy model.

### 3 Building a resource for Italian Super-sense Tagging

#### 3.1 Italian Syntactic-Semantic Treebank

The resource to train the SST tagger was built starting from the Italian Syntactic-Semantic Treebank (ISST), one of the main outcomes of an Italian national project, SI-TAL (Montemagni et al., 2003). ISST is a multi-layered corpus, annotated at the orthographic, morpho-syntactic, syntactic and lexico-semantic levels. It is partitioned into two different sections, a “balanced” corpus, testifying general language usage and a specialized corpus with texts belonging to the financial domain.

Balanced part	215,606
Specialized Part	89,941
<b>Overall size</b>	<b>305,547</b>

Table 4: ISST Corpus statistics

In ISST, lexico-semantic annotation consists in the assignment of semantic tags to content words or to sequences of words corresponding to a single unit of sense (e.g. compounds, idioms). In particular, annotation was restricted to the classes of nouns, verbs and adjectives and corresponding multi-word expressions. The annotation was carried out in lexicographic mode (as opposed to full-text mode), i.e. it was carried out on a lemma-by-lemma basis and was restricted to selected lemmas: the final resource contains 81,236 content words annotated at the lexico-semantic level.

ISST semantic tags convey two main types of information:

- *sense* of the target word(s) in the specific context: ItalWordNet (IWN) was the reference lexical resource used for the sense tagging task (Roventini et al., 2003). It is organized into two different modules, a general one and a specialized one with financial and computational terminology;
- other types of lexico-semantic information not included in the reference lexical resource, e.g. for marking of figurative usages or named entity categorization.

#### 3.2 Adding super-senses to ISST

For the specific concerns of the work reported in this paper, we used the whole ISST corpus, focusing on the morpho-syntactic and lexico-semantic annotation levels.

In particular, the following information types were extracted from the ISST resource:

- part of speech, lemma, and morpho-syntactic features (such as number, person, gender, etc.) from the morpho-syntactic annotation level;
- IWN senses and named entities categorization (in classes like human entity, artifact, institution, location, etc.) from the lexico-semantic level.

In the original ISST corpus, content words are partially tagged with  $\langle lemma, sense \rangle$  pairs that allow, when present, to map each token to an ItalWordNet synset. To turn ISST into a resource for super-sense tagging, we performed, where possible, a semi-automatic mapping of IWN synsets into WordNet super-senses. This was done exploiting the taxonomical organization of IWN word senses and the Inter-Lingual Index (ILI), linking Italian synsets to the WordNet corresponding synsets (Figure 1).

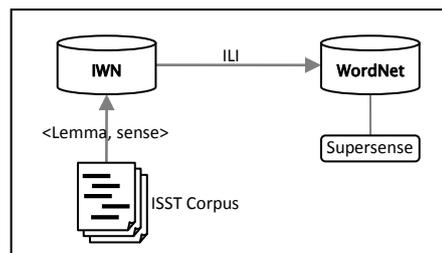


Figure 1: ISST super-sense mapping

The resulting corpus, annotated with super-senses, will be henceforth referred to as ISST-SST.

Since the ILI covers only a subset of the IWN senses, the automatic phase of the mapping included the following steps (Figure 2):

1. starting from each annotated Italian synset  $s_i$  in ISST:
  - if  $s_i$  was part of the ILI, we directly reach the linked synset  $s_e$  in WordNet;
  - if  $s_i$  was not part of the ILI, we reached its first hyperonym in the IWN hierarchy connected through an ILI to a WordNet synset  $s_e$ ;
2. the WordNet super-sense associated to  $s_e$  was then assigned as an attribute to  $s_i$  in the ISST-SST corpus.

This way, each original Italian synset had a chance of



extractor is applied and the classifier computes a probability distribution for the tags to assign to the current token.

The corpora used for training the tagger use the IOB2 annotation convention (Inside, Outside, Begin) as shown in Listing 2. The chunker has an option to split the IOB tags into a more refined set, which performed well for NER but proved to be less relevant for SST.

Commosa	Vpsfs	B-verb.emotion
,	FF	O
Fiona	SP	B-noun.person
May	SP	I-noun.person
ha	VAip3s	O
parlato	Vpsms	B-verb.communication
,	FF	O
tramite	E	O
(...)		

Listing 2: Example of IOB annotation

Maximum Entropy is effective for chunking since it does not assume independence of features and is typically more accurate than an Average Perceptron classifier.

#### 4.1.1 Maximum Entropy and Dynamic Programming

The Maximum Entropy framework estimates probabilities based on the principle of making as few assumptions as possible, other than the constraints imposed. Such constraints are derived from training data, expressing some relationship between features and outcome. The probability distribution that satisfies the above property is the one with the highest entropy. It is unique, agrees with the maximum-likelihood distribution, and has the exponential form (Della Pietra et al., 1997):

$$p(o|h) = \frac{1}{Z(h)} \prod_{j=1}^k \alpha_j^{f_j(h,o)},$$

where  $o$  refers to the outcome,  $h$  the history or context, and  $Z(h)$  is a normalization function. The features used in the Maximum Entropy framework are binary. An example of a feature function is:

$$f_j(h,o) = \begin{cases} 1 & \text{if } o = B - \text{noun.location}, \text{FORM} = \text{Washington} \\ 0 & \text{otherwise} \end{cases}$$

The parameters  $\alpha_j$  are estimated by a procedure called Generalized Interactive Scaling (GIS) (Darroch & Ratcliff, 1972). This is an iterative procedure that improves the estimation of parameters at each iteration. Since the Maximum Entropy classifiers assigns tags to each token independently, it may produce inadmissible sequences of tags. Hence a dynamic programming technique is applied to select correct sequences. A probability is assigned to a sequence of tags  $t_1, t_2, \dots, t_n$  for sentence  $s$ , based on the probability of the transition between two consecutive tags  $P(t_{i+1} | t_i)$ , and the probability of a tag  $P(t_i | s)$ , obtained from the probability distribution computed by Maximum Entropy:

$$P(t_1, t_2, \dots, t_n) = \prod_{i=1}^n P(t_i | s) P(t_i | t_{i-1})$$

In principle the algorithm should compute the sequence with maximum probability.

We use instead a dynamic programming solution which operates on a window of size  $w = 5$ , long enough for most super-senses. For each position  $n$ , we compute the best probability  $PB(t_n)$  considering the  $n$ -grams of length  $k < w$  preceding  $t_n$ :

$$PB(t_n) = \max_k PB(t_{n-k-1}) \dots PB(t_{n-1})$$

A baseline is computed first, assuming that the  $k$ -gram is made all of 'O' (outside) tags:

$$PB_O(t_n) = \max_k PB(t_{n-k-1}) P(t_{n-k} = O) \dots P(t_{n-1} = O)$$

Similarly for each class  $C$  we compute:

$$PB_C(t_n) = \max_k PB(t_{n-k-1}) P(t_{n-k} = C) \dots P(t_{n-1} = C)$$

and finally:

$$PB(t_n) = \max(PB_O(t_n), \max_C PB_C(t_n))$$

## 4.2 Features extraction

The modular architecture of the chunker relies on the use of the abstract class `FeatureExtractor` for extracting features during training and analysis. The class `SstFeatureExtractor` is defined as a specialization of that abstract class, specifically designed for Super-sense tagging. It extracts a basic set of features from the current and surrounding tokens. More specific features, necessary for a given task or for a given language, can be specified in a configuration file.

There are two mechanisms to specify the additional features to extract: as attributes of the tokens or as token features expressed by a regular expression.

An example of an *attribute feature* is the following:

```
POSTAG -1 0
```

which requests to use as features the POS tag of the previous (-1) and current (0) token.

*Token features* can be expressed with regular expressions, for instance, in:

```
MorphFeature FORM ^\p{Lu} -1 +1
MorphFeature FORM ^\p{Lu}*$ 0
```

The first line indicates to use as features the property of starting with an uppercase letter (Unicode property `Lu`) for the previous (-1) and next token (+1); the second line indicates the feature representing that the current token consists of all upper case letters.

## 4.3 Features specification

In this section we describe the features we used to train the SST. They can be divided into three main categories:

- attributes features;
- local features;

- global features.

It is worth noting that the tagger does not use any external resource, such as dictionaries or gazetteers. Besides, we could not use the first sense heuristic used in Ciaramita and Altun (2006), since the Italian WordNets do not provide sense frequency statistics.

#### 4.3.1 Attributes Features

Attributes features are related to the position of the attributes surrounding the current token. After various tuning experiments, we obtained the best results with this combination:

POSTAG	-2 -1 0 1 2
CPOSTAG	-1 0

where POSTAG contains a fine-grained Part-of-Speech with morphological information and CPOSTAG is the coarse-grained POS tag.

#### 4.3.2 Local features

Local features are morphological features extracted from the analysis of the current word and the context in which it appears. There are two kinds of local features:

- *Features of Current Word*: first word of sentence and capitalized; first word of sentence and not capitalized; two parts joined by a hyphen.
- *Features from Surrounding Words*: both previous, current and following words are capitalized; both current and following words are capitalized; both current and previous words are capitalized; word is in a sequence within quotes.

#### 4.3.3 Global features

Global features are properties holding at the document level. For instance, if a word in a document was previously annotated with a certain tag, then it is likely that other occurrences of the same word should be tagged similarly. Global features represent these properties. They are particularly useful in cases where the word context is ambiguous but the word appeared previously in a simpler context.

## 5 Experiments and results

Using the new resource and the SuperSense Tagger based on Maximum Entropy model, presented in chapter 4, we obtain the following results:

SuperSense	%	Precision	Recall	F1
adj.all	12,39%	81.78	91.92	86.56
adj.pert	3,84%	84.19	73.68	78.59
adv.all	5,69%	95.80	98.45	97.10
noun.Tops	11,20%	86.46	66.94	75.45
noun.act	0,02%	78.49	88.91	83.38
noun.animal	2,47%	100.00	40.00	57.14
noun.artifact	4,88%	69.04	69.42	69.23
noun.attribute	0,09%	85.77	70.16	77.19
noun.body	3,86%	75.00	100.00	85.71
noun.cognition	7,32%	83.59	71.13	76.86

noun.communication	1,33%	79.21	82.86	80.99
noun.event	0,08%	82.95	67.72	74.56
noun.feeling	0,28%	60.00	75.00	66.67
noun.food	6,09%	100.00	13.16	23.26
noun.group	2,97%	69.64	51.92	59.49
noun.location	0,14%	75.60	72.52	74.03
noun.motive	0,28%	100.00	50.00	66.67
noun.object	3,61%	97.87	55.42	70.77
noun.person	0,32%	60.92	86.65	71.54
noun.phenomenon	0,16%	100.00	78.95	88.24
noun.plant	10,11%	100.00	13.64	24.00
noun.possession	0,09%	90.75	79.54	84.78
noun.process	0,61%	100.00	57.14	72.73
noun.quantity	0,47%	74.11	77.57	75.80
noun.relation	0,02%	86.67	59.09	70.27
noun.shape	1,34%	0.00	0.00	0.00
noun.state	0,38%	88.94	77.08	82.59
noun.substance	2,03%	86.11	58.49	69.66
noun.time	0,83%	88.19	95.22	91.57
verb.body	0,05%	100.00	21.21	35.00
verb.change	2,65%	70.41	71.90	71.15
verb.cognition	1,66%	86.78	68.33	76.46
verb.communication	3,49%	76.64	87.38	81.66
verb.competition	0,17%	44.44	28.57	34.78
verb.consumption	0,01%	93.18	69.49	79.61
verb.contact	0,33%	64.66	64.66	64.66
verb.creation	0,75%	70.99	59.62	64.81
verb.emotion	0,15%	69.23	45.00	54.55
verb.motion	0,81%	55.21	43.44	48.62
verb.perception	0,69%	85.88	54.48	66.67
verb.possession	1,12%	92.97	65.75	77.02
verb.social	1,53%	57.06	59.51	58.26
verb.stative	3,67%	71.15	83.93	77.02
verb.weather	0,01%	0.00	0.00	0.00
<b>ALL</b>	<b>100%</b>	<b>79.92</b>	<b>78.30</b>	<b>79.10</b>

Table 7: Best results on ISST-SST

The F1 score of 79.10 represents a value higher than the current state of the art on the English language as shown in Table 8.

	Precision	Recall	F1
<b>English</b> Ciaramita et al.	76,65	77,71	77,18
<b>Italian</b> <i>our</i>	79.92	78.30	79.10

Table 8: State of art for English and Italian

This improvement in performance is due in particular to the following contributions:

- the new corpus: about +4.5 on the F1 score;
- the different algorithm and the tuning of features: about +10 on the F1 score.

Corpus contribution has been computed considering the difference in performance between MultiSemCor and ISST-SST, using our SST, on the same test set.

Instead, the improvement due to the adoption of Maximum Entropy instead of Average Perceptron algorithm was calculated using the SST developed by Ciaramita & Altun, and our SST on the ISST-SST corpus. The results obtained in this experiment are summarized in Table 9.

	Precision	Recall	F1
SST Ciaramita et al.	71.30	67.89	69.56
SST Maximum Entropy	79.92	78.30	79.10

Table 9: Average Perceptron SST vs Maximum Entropy SST on ISST-SST

The efficiency in tagging is also competitive: the real time for tagging a corpus of 300,000 tokens is about 4 seconds with our tagger and 66 seconds with the average perceptron classifier.

## 6 Applications

The tagger has already been integrated in the TANL pipeline of linguistic tools of the SemaWiki project, and is currently being used in a Deep Search application on the Italian Wikipedia.

An interesting example of query to Deep Search that uses super-senses is the following:

```
deprel [DEP/subj:Edison WNSS/creation:*]
```

It asks to find all the sentences within Wikipedia where the subject is Edison and the verb is any one among those classified under the super-sense `verb.creation`. Results are therefore related to what Edison invented, what he built, what he created and so on.

## 7 Conclusions

We have achieved state-of-the-art performance on SST tagging, by developing an SST corpus and a tagger for Italian texts. The improvement in performance of SST can also be regarded as an indirect proof of the good quality of the ISST-SST resource. An accurate and efficient SST is useful for many tasks of semantic analysis and can be used as input to other analysis tools such as WSD, parsers and semantic role labelers. In the future we plan to further improve the ISST-SST corpus and exploit the tagger in question answering. This will give us more insights on the practical value of annotating texts with the 45 super-senses of WordNet.

## 8 Acknowledgments

The SemaWiki project was partially funded by the Fondazione Cassa di Risparmio di Pisa. We thank Massimiliano Ciaramita and Davide Picca for helping us in the reconstruction of their previously published results on super-sense tagging for Italian.

## 9 References

- G. Attardi et al. (2009) Tanl (Text Analytics and Natural Language Processing). SemaWiki project: <http://medialab.di.unipi.it/wiki/SemaWiki>.
- G. Attardi, S. Dei Rossi and M. Simi (2010). The Tanl Pipeline. Proc. of Workshop on Web Services and Processing Pipelines in HLT, co-located LREC 2010, Malta.
- G. Attardi, A. Fuschetto, F. Tamperi, M. Simi and E. M. Vecchi (2009). Experiments in tagger combination: arbitrating, guessing, correcting, suggesting. Evalita 2009, Reggio Emilia.
- L. Bentivogli, P. Forner, and E. Pianta (2004). Evaluating Cross-Language Annotation Transfer in the MultiSemCor Corpus. Proceedings of Coling 2004, Geneva.
- H.L. Chieu, H.T. Ng (2003). Named Entity Recognition with a Maximum Entropy Approach. Proceedings of CoNLL-2003, pp. 160-163. Edmonton, Canada.
- M. Ciaramita, Altun, Y. (2006). Broad-Coverage Sense Disambiguation and Information Extraction with a Supersense Sequence Tagger. EMNLP.
- M. Ciaramita and M. Johnson (2003). Supersense tagging of unknown nouns in WordNet. Proc. of EMNLP 2003.
- M. Collins (2002). Discriminative Training Methods for Hidden Markov Models: Theory and Experiments with the Perceptron Algorithm.: Proceedings of EMNLP.
- J. N. Darroch and D. Ratcliff (1972). Generalized Iterative Scaling for Log-Linear Models. *Annals of Mathematical Statistics*, 43(5), pp. 1470–1480.
- S. Della Pietra, V. Della Pietra, and J. Lafferty (1997). Inducing Features of Random Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(4), pp. 380–393.
- C. Fellbaum (Ed.) (1998). *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge.
- P. Halácsy, A. Kornai, and C. Oravecz (2007). HunPos: an open source trigram tagger. Proceedings of the ACL 2007, Prague.
- S. Montemagni, F. Barsotti, M. Battista, N. Calzolari, O. Corazzari, A. Lenci, A. Zampolli, F. Fanciulli, M. Massetani, R. Raffaelli, R. Basili, M. T. Pazienza, D. Saracino, F. Zanzotto, N. Mana, F. Pianesi, R. Del monte (2003). Building the Italian Syntactic-Semantic Treebank. In Anne Abeillé, Building and using Parsed Corpora, Language and Speech series, Kluwer, Dordrecht.
- G. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. Miller (1993). Introduction to WordNet: An on-line lexical database. Technical report, Princeton.
- D. Picca, A. Gliozzo, M. Ciaramita (2008). SuperSense Tagger for Italian. Proc. of LREC 2008, Marrakech.
- A. Roventini, A. Alonge, N. Calzolari, B. Magnini, F. Bertagna (2000). ItalWordNet: a Large Semantic Database for Italian. Proceedings LREC 2000, Athens.