Computational Models of Event Type Classification in Context

Alessandra Zarcone[1,2] **Alessandro Lenci**[1]

[1]Università di Pisa - Dipartimento di Linguistica "T. Bolelli"
Via S. Maria 36 - 56126 - Pisa, Italy

[2]Scuola Normale Superiore, Laboratorio di Linguistica
Piazza dei Cavalieri 7 - 56126 - Pisa, Italy

a.zarcone@sns.it, alessandro.lenci@ilc.cnr.it

Abstract

Verb lexical semantic properties are only one of the factors that contribute to the determination of the event type expressed by a sentence, which is instead the result of a complex interplay between the verb meaning and its linguistic context. We report on two computational models for the automatic identification of event type in Italian. Both models use linguistically-motivated features extracted from Italian corpora. The main goal of our experiments is to evaluate the contribution of different types of linguistic indicators to identify the event type of a sentence, as well as to model various cases of context-driven event type shift. In the first model, event type identification has been modelled as a supervised classification task, performed with Maximum Entropy classifiers. In the second model, Self-Organizing Maps have been used to define and identify event types in an unsupervised way. The interaction of various contextual factors in determining the event type expressed by a sentence makes event type identification a highly challenging task. Computational models can help us to shed new light on the real structure of event type classes as well as to gain a better understanding of context-driven semantic shifts.

1. Introduction

The *event type* (aka *Aktionsart*) expressed by a predicate is a crucial component of the sentence temporal constitution. By event type we refer here to the standard Vendler's classification of predicates into *state* (STA), *activity* (ACT), *accomplishment* (ACC) and *achievement* (ACH) (Vendler, 1967). These four categories are further cross-classified with respect to the dimensions of telicity and durativity. Accomplishment and achievement predicates are *telic*, while states and activities are *atelic*. On the other hand, states, activities and accomplishments are *durative*, and achievements are *non durative* (see table 1).

Event type	[telic]	[durative]	[dynamic]
STA	−	+	−
ACT	−	+	+
ACC	+	+	+
ACH	+	−	+

Table 1: The features of Vendler's event types

Semantic literature reports well-known linguistic diagnostics, typically used to classify the event type of a predicate (Dowty, 1979; Bertinetto, 1986; Pustejovsky, 1995; Rothstein, 2004).

Although the event type is often referred to as "lexical aspect", the semantic features of the verb are only one of the factors contributing to determine the event type expressed by a sentence, which is instead the result of a complex interplay between the verb meaning and its linguistic context. Actually, various contextual factors (such as arguments, their definiteness, their animacy, temporal adverbials, verb's morphology, etc.) can *shift* the verb event type to a new class.

Consider the following examples:

1. (a) *John has been reading for the whole day* (atelic, durative)
   (b) *John has read "The Great Gatsby" in an hour* (telic, durative)
   (c) *John has been reading papers for the whole day* (atelic, durative)
2. (a) *John has been pushing the chart* (atelic, durative)
   (b) *John has pushed the chart to the checkout line* (telic, durative)
3. (a) *The train arrived at 5 o'clock* (telic, non durative)
   (b) *Europeans had been arriving sporadically, sometimes with long intervals between arrivals* (atelic, durative)
4. (a) *The path goes from the street into the forest* (state)
   (b) *The cat is going to the door* (dynamic)
5. (a) *John has hung the picture on the wall* (telic, dynamic)
   (b) *The picture hangs on the wall* (state)

Read has an atelic, durative meaning in (1a). However, a definite direct object can turn this predicate into a telic event, e.g. (1b). Conversely, when the same verb occurs with a bare plural object the event is again an atelic one (1c). Differently from the case of *read*, the definite direct object is not enough to turn *push* into a telic event in (2a). This verb can also be turned into a telic event if we add a PP expressing the destination of the movement (2b). Likewise, even a prototypical achievement predicate like *arrive* (3a)

Verbs	OCCURR.	STA	ACT	ACC	ACH
28	3129	583	430	822	1294

Table 2: The composition of the corpus for the supervised model

Verb	Translation
arrivare	to arrive
capire	to comprehend / realize / understand
chiamare	to call
chiedere	to ask
chiudere	to close
comprendere	to comprehend / realize / understand
conoscere	to know / get to know
controllare	to check
costituire	to constiture / establish
entrare	to enter / go in
indicare	to point
lasciare	to leave / let
lavorare	to work
mettere	to put
morire	to die
parlare	to speak / to talk
partire	to depart / leave
passare	to pass / spend (time)
portare	to bring / carry
prendere	to take
presentare	to introduce / to present
scrivere	to write
spiegare	to explain
tornare	to come back
trattare	to deal / transact / process
trovare	to find
vendere	to sell
vincere	to win

Table 3: The verbs in the corpus for the supervised model

can be turned into a durative event by a plural subject (3b). Other factors possibly affecting the sentence event type are the subject's animacy or the verb's morphology (4a vs. 4b, 5a vs. 5b).
Verbs like those shown in the examples, are usually referred to as *actionally polysemous* or *hybrid* (Bertinetto, 1986), because they can exhibit different event type values in different contexts. Context-driven *event type shifts* raise the question of how the sentence event type can be recognized, and how the contribution of different aspects of the linguistic context to its identification can be properly modelled. Moreover, contextual features do not act as necessary and sufficient conditions in determining a certain event type. In fact, for each event type class it is impossible to define a set of linguistic contextual elements univocally associated with it.

2. Goal of the paper

Though event types lie at the centre of a long tradition of research in formal semantics, their computational modelling has received little attention. Notable exceptions are Siegel and McKeown (2000) and Palmer et al. (2007). However, Siegel and McKeown (2000) do not deal with the problem of context-driven event type shifts. They rather use different types of machine learning methods to recognize what they call "the fundamental aspectual category". Moreover, they only train binary classifiers to distinguish states from events, and culminated (telic) from non culminated (atelic) clauses. On the other hand, Palmer et al. (2007) do not specifically focus their automatic classification experiments on Aktionsart, and are instead more concerned with a wider notion of "situation type", encompassing also speech-act types, abstract entities (e.g. facts, propositions), generics, etc.
In this paper, we report on two computational models for the automatic identification of event type in Italian. We assume that the event type expressed by a clause is determined by the complex interaction among different features, such as the verb's arguments, their definiteness, tense-aspectual morphology, adverbials, etc. Our models use linguistically-motivated features extracted from Italian corpora to evaluate the neat contribution of different types of linguistic indicators to the event type identification task, as well as to model various cases of context-driven event type shift.
Section 3. reports on the first model, that formalizes event type identification as a supervised classification task performed with Maximum Entropy classifiers (MaxEnt; Berger et al. (1996)). Section 4. reports on the second model, which uses Self-Organizing Maps (SOMs; Kohonen (1997)) to define and identify event types in an unsupervised way.

3. The supervised model (MaxEnt)

3.1. Training data

No corpora annotated with event type information were available for Italian. Therefore, 3129 occurrences of 28 Italian verb predicates from the Italian Syntactic-Semantic Treebank Montemagni et al. (2003) were manually annotated by one of the authors with their event type. We used 4 classes: *state*, *process*, *achievement* and *accomplishment*. Table 2 shows the distributions of the verb occurrences in the training set with respect to the four event types. It is worth remarking that the class assigned to each verb token corresponds to its contextually determined event type. In fact, event type assignment was decided on the ground of the whole set of linguistic features available at the sentence level. Therefore, the same verb type may be associated with different event types in the training corpus.
Since Italian verb predicates can vary with respect to their degree of hybridism, the corpus was further divided according to the verb predicates ambiguity.

60% group - it is the most polysemous group and contains the verbs whose most frequent event type covers less than 60% of their tokens;

70% group - it includes the verbs of the previous group, plus those verbs whose most frequent event type covers less than 70% of their tokens;

80% group - it includes the verbs of the previous group, plus those verbs whose most frequent event type covers less than 80% of their tokens;

90% group - it includes the verbs of the previous group, plus those verbs whose most frequent event type covers less than 90% of their tokens.

3.2. The MaxEnt model

In the first model, event type classification has been performed with Maximum Entropy classifiers (Berger et al., 1996), trained on a corpus annotated with the proper event type of the predicate. Given a linguistic context c, and a category $a \in A$ dependent on c, the conditional probability $p(a|c)$ is found assuming that the distributions of a set of relevant features $f_i(a,c)$ of c are the only probabilistic constraints involved (whose distributions are learned from the training corpus). It can be proved that the only probability distribution p that is coherent with this assumption is the one with the maximum entropy, that is (Berger et al., 1996):

$$p\left(a \mid c\right) = \frac{1}{Z_c} \prod_{i=1}^{k} {a_i}^{f_i(a,c)}$$

where $f_i(a,c)$ are values of k features of (a,c), $a_1...a_k$ are the features weights and Z_c is a normalization factor.
In the training phase, feature weights were estimated by using the GIS (Generalized Iterative Scaling) algorithm, in its AMIS software implementation (Miyao, 2002). During the test phase, for each new context c the model combines the estimated weights to compute $p(a|c)$ for every $a \in A$. The category a, given the context c, is the one with the highest probability ($argmax\left(p\left(a \mid c\right)\right)$).
Experiments were carried out with a 10-fold cross-validation method applied both on the whole corpus and on each of the polysemy grous defined above. As evaluation measures we used:

- *accuracy*, i.e. the percentage of correctly classified verbs occurrences. The baseline for the whole corpus and for each polysemy group was calculated by assigning to every verb its most frequent event type in the corpus (group).

- *precision* (P) and *recall* (R) for each event type, which were then combined into the *f-measure* ($2PR/P+R$).

3.3. Feature selection

Both models use linguistically-motivated features extracted from Italian corpora. These features, which are very well-known in the linguistic literature for being (positively or negatively) correlated with particular event types (Dowty, 1979; Bertinetto, 1986; Pustejovsky, 1995; Rothstein, 2004), include:

adverbial features - in the literature they are among the main "event type" diagnostics, but they are not very frequent in corpora data. They include various types of adverbials as:

- temporal adverbs ("in X time", "for X time", etc.);
- intentional adverbs ("deliberately", "intentionally", etc.);
- frequency adverbs ("rarely", "often", etc.);
- iterative adverbs ("X times");

morphological features - although actionality and aspect are independent categories, it is possible to observe typical correlations between some event types and specific aspectual values (Comrie, 1976). This set of features includes verb morphological tense-aspectual values, such as:

- present tense;
- imperfect tense;
- future tense;
- simple past;
- perfect tenses;
- progressive periphrasis;

syntactic and argument structure features - they include verb morphosyntactic, syntactic and semantic features of verb arguments, which are typically held responsible for event type shifts (see examples 1a-5b in section 1):

- absence of arguments besides the subject;
- presence of direct object;
- presence of indirect object;
- presence of a locative argument;
- presence of a complement sentence;
- passive diatesis;
- subject and direct object, number, animacy and definiteness.

All the features were extracted from the corpus in a semi-automatic way.

3.4. The experiments

We tested MaxEnt in different types of experiments. In the first one, we used the whole set of features, while in second one we trained the classifiers only using specific subsets of linguistic cues. While in the both these cases, MaxEnt perfomed 4-way classifications, in the last experiment it was applied to carry-out 2-way classifications with respect to the three defining features of event types: *telicity*, *durativity*, and *dynamicity* (cf. Section 1.).

3.4.1. Experiment 1: complete feature set

The first model was built by using the whole set of features. See the results in table 4 and 5. The model is able to outperform the baseline, showing that contextual features play an important role in event type classification and can therefore be interpreted as statistic cues to guide this task.
A deeper analysis of the system mistakes has revealed that they mostly concern cases in which the verb either appears in a non-finite clause (i.e. infinitive, participle and gerund structures) or is used with an idiomatic sense. Consider the following examples from our corpus:

- Umberto Eco ha potuto divertirsi a *prendere* un po' tutti *per il bavero*.
  Umberto Eco has been able to pull everybody's sleeve.

- Questa è una fetta di Croazia dove fino a ieri sera noi non potevamo neppure *mettere piede*.
  This is a part of Croatia where up to yesterday evening we weren't even allowed to set foot in.

Sentences like these represent very hard cases for our model, because their event type meaning is completely idiosyncratic.

	Baseline	Exp 1
60 % group	56.1%	69.3%
70 % group	60%	72.8%
80 % group	64.6%	75.5%
90 % group	69.6%	78.4%
Whole corpus	79.8%	85.4%

Table 4: model accuracy

	ACT	STA	ACC	ACH
60 % group				
precision:	0.33	0.75	0.64	0.73
recall:	0.2	0.69	0.85	0.69
f-measure:	0.25	0.72	0.73	0.71
70 % group				
precision:	0.33	0.76	0.72	0.73
recall:	0.15	0.73	0.83	0.73
f-measure:	0.21	0.75	0.77	0.73
80 % group				
precision:	0.51	0.79	0.79	0.74
recall:	0.39	0.71	0.87	0.73
f-measure:	0.44	0.75	0.83	0.74
90 % group				
precision:	0.54	0.84	0.79	0.78
recall:	0.42	0.8	0.86	0.79
f-measure:	0.47	0.82	0.82	0.79
whole corpus				
precision:	0.84	0.83	0.84	0.88
recall:	0.74	0.78	0.89	0.9
f-measure:	0.79	0.8	0.86	0.89

Table 5: Precision and recall results from experiment 1

	ACT	STA	ACC	ACH
adverbial features				
precision:	0.49	0.35	0.29	0.66
recall:	0.05	0.1	0	0.14
f-measure:	0.09	0.15	0	0.24
morphological features				
precision:	0.36	0.38	0.15	0.53
recall:	0.08	0.62	0	0.49
f-measure:	0.13	0.47	0	0.51
syntactic and argument structure features				
precision:	0.89	0.79	0.78	0.86
recall:	0.66	0.7	0.92	0.88
f-measure:	0.76	0.75	0.84	0.87
whole set of features				
precision:	0.84	0.83	0.84	0.88
recall:	0.74	0.78	0.89	0.9
f-measure:	0.79	0.8	0.86	0.89

Table 6: Precision and recall results from experiment 2

3.4.2. Experiment 2: feature subsets

The aim of the second battery of experiments is to show the contribution to event type classification offered by feature subsets corresponding to specific types of linguistic information (e.g. morphology, temporal adverbs, etc.). Precision and recall values are reported in table 6.
Adverbial features are very good in providing high-precision event type classifications (they are particularly useful to identify activities and accomplishments), but recall values are low because those features are very sparse. In fact, they appear just in 16% of the sentences in the training corpus. Morphological features are much more frequent, but less precise. They mostly help to identify states and achievements. Conversely, the distinction between activities and accomplishments improves only when syntactic features are added to the model, significantly raising precision and recall values.
Note that the highest precision and recall values are nevertheless those obtained in experiment 1, in which we used the complete feature set was used. This proves that no specific type or level of linguistic information is singularly able to determine the event type of a sentence, but it is rather the complex interaction of different linguistic cluses that can achieve the optimal level of event type discrimination.

3.4.3. Experiment 3: 2-way classifications

In experiment 3 we built 3 different models, and we trained each of them to perform a 2-way classification. Instead of using the 4 categories we had been using so far, we used the distinctive features in table 1: the first model was trained to distinguish *durative* event types from *non-durative* ones, the second one to distinguish *dynamic* event types from *non-dynamic* ones, and the third one to distinguish *telic* event types from *non-telic* ones. This experiment was performed with the model trained with the complete feature set.
A 2-way classification is of course an easier task, and consistently the baseline is higher (see table 7 for baseline and accuracy values). Nevertheless, the system outperforms the

baseline in every polysemy group and in the whole corpus as well. It is worth observing that MaxEnt more easily distinguish between dynamic and non-dynamic events, and then between telic and non-telic ones. Conversely, durativity appears to be the hardest feature to discriminate.

	Baseline	Exp 3
$+/-$ DUR		
60 % group	63.9%	72.8%
70 % group	68.3%	74.3%
80 % group	75.5%	79.1%
Whole corpus	88.3%	90.6%
$+/-$ DIN		
60 % group	60.9%	79.9%
70 % group	62.2%	84.9%
80 % group	70%	85.4%
Whole corpus	87.7%	92%
$+/-$ TEL		
60 % group	56.9%	79.5%
70 % group	66.8%	81.7%
80 % group	71.9%	83.2%
Whole corpus	84.4%	89.9%

Table 7: model accuracy in 2-way classification

3.5. Establishing an upper-bound for event type identification

One could argue how an upper-bound for the performance of our systems can be defined. Actually, event type identification seems to be a highly challenging task even for humans.

In order to investigate this issue, we randomly selected 100 sentences from the annotated corpus, containing verbs from the 70% polysemy group. Three subjects were asked to tag each sentence by choosing only one class out of the four event types used in the MaxEnt experiments. All taggers were trained linguists, with a long experience with actionality. There was no temporal limit for the annotation. Table 8 reports the accuracy of the three annotators (T1, T2, T3) – together with the score achieved by the MaxEnt model – calculated with respect to the gold standard represented by the annotated corpus we introduced in Section 3.1.. Not only is the human accuracy well below 100%, but that it is also totally comparable with the MaxEnt model performances. These results show the inherent complexity of event type identification, especially when applied to real corpus data, which are far from being as clear-cut as the standard examples typically reported in the literature.

	Accuracy
T1	73%
T2	44%
T3	67%
MaxEnt	76%

Table 8: Results of the human tagging experiment

4. The unsupervised model (SOM)

4.1. Data for the unsupervised model

A sample of 40 Italian verbs have been selected for their high degree of prototipicality with respect to the four event types in table 1 (10 verbs for each category). Following the approach in Lagus and Airola (2005), every verb has been represented as a distributional vector, recording its co-occurrence frequency with a certain number of context features. Consistently with the "distributional hypothesis" (Harris, 1968), we assume that two verbs have similar Aktionsart values if they have similar context feature distributions. 28 features correspond to those selected for the MaxEnt model, because of their high correlation with event types. We added 40 extra features to encode the lexical entry of each verb. Thus, the verb vectors are orthogonal with respect to the 40 extra features.

Feature frequency was estimated from "La Repubblica" (Baroni et al., 2004), a large corpus of Italian newspaper texts of about 450 million tokens. The corpus was automatically PoS tagged, lemmatized and shallow-parsed. Vectors were weighted using the following global weight function (Dumais, 1990), so that features more uniformly distributed among the verbs have a smaller weight:

$$w_f = 1 - \sum_v \frac{p_{fv} \log_2(p_{fv})}{\log_2(\#totv)}$$

$$p_{fv} = p(v|f) = \frac{\#f_v}{\#f}$$

where $\#totv$ is the number of vectors and p_{fv} is the probability to observe a given vector v and a given feature f together. All the vectors have been weighted and then normalized before every experiment.

4.2. The SOM model

Self-Organizing Maps (SOM; Kohonen (1997)) are a particular kind of unsupervised neural network, used to project n-dimensional vectors into a 2-dimensional space (map) preserving the topological properties of the input space. Their tipical architecture is composed of a honeycomb network of n nodes. Every node represents a k-dimensional vector (where $k > n$), and the input nodes are randomly inizialized. When the map receives an input vector, it activates in parallel all its nodes: node activation is proportional to its similarity with the input vector. The node with the highest activation (*best matching unit*) is selected, and its vector – together with the ones of the units close to it – is modified to become more similar to the input vector.

4.3. The experiments

4.3.1. Experiment 4: building a SOM for event type

The SOM we trained with the first 40 verbs was a honeycomb map of 100 nodes and it was developed using the software Matlab 7.3.0 with the SOM package. The map is shown in figure 1.

STA and ACT categories look well defined and distinct from ACH and ACC. Not surprisingly, ACH and ACC cover the same area of the map. Recall from table 1 that these two categories differ just for the *durative* feature, that appears to be the hardest one for event type identification,

Verb	Translation
ascoltare	to listen
aggiustare	to fix
amare	to love
appartenere	to belong
aprire	to open
arrivare	to arrive
cadere	to fall
camminare	to walk
cantare	to sing
credere	to believe
cucinare	to cook
dimostrare	to prove
disegnare	to draw
dormire	to sleep
elaborare	to compute / work out
lavorare	to work
leggere	to read
lottare	to struggle
mancare	to miss
morire	to die
navigare	to sail
parlare	to speak / talk
partire	to depart / leave
piangere	to cry
possedere	to own
preparare	to prepare
pulire	to clean
raggiungere	to reach
risiedere	to reside
risolvere	to solve
ritenere	to deem / reckon / retain
rompere	to break
sapere	to know
scoprire	to find out
scrivere	to write
sembrare	to appear / look / seem
temere	to fear
tracciare	to draw / trace
vendere	to sell
vincere	to win

Table 9: The verbs in the corpus for the unsupervised model

mainly because it often depends on the pragmatic context rather than on overt linguistic clues.

4.3.2. Experiment 5: an IR-like model

In order to go beyond the qualitative evaluation of the SOM built in Experiment 4, we have tried to evaluate our unsupervised model for event type classification in analogy with the vector space model in Information Retrieval.
In the training phase, the SOM has been trained with the 40 verbs of Experiment 4. Recall that those 40 vectors describe the behaviour of 40 verbs in all the context they have been found. Then, the SOM has been used to model context-driven event type shifts. Test items are very sparse vector representing verbs specific *context types*. In each vector there is a small number of active dimensions, cor-

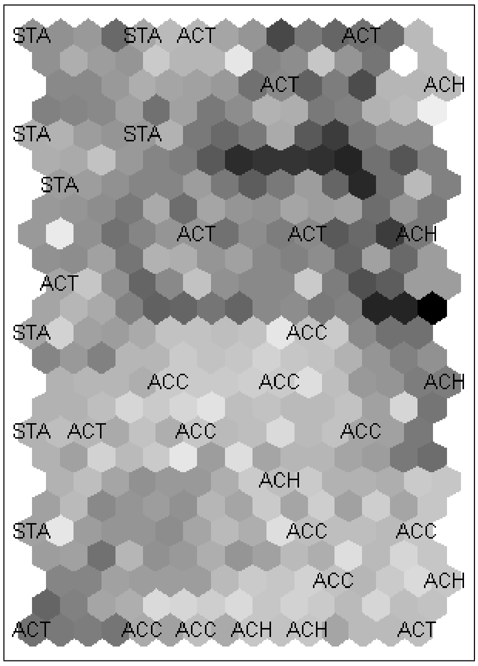

Figure 1: Experiment 4: the SOM

responding to the linguistic features available in a certain context. When specific context features shift the event type value of a verb to a new class (e.g. by turning an activity into an accomplishment), we expect this verb to change its position in the "Aktionsart semantic space", getting near to the cluster of verbs belonging to the new class. Therefore, every vector in the test set has been used as a sort of "query vector" and given in input to the verb semantic map to identify the verb cluster in the SOM corresponding to the best-matching unit activated by the query vector. If the best-matching unit for the query vector was labelled in the training phase, the answer for the query vector was the category of that label. Otherwise we have selected the category of the nearest labelled unit.
This experiment was carried out with a test set of 40 new contexts, which were weighted with the same feature weighting scheme we used with the training set. Results are shown in table 9. Not surprisingly, a large part of the mistakes is found in discriminating between the two telic categories (ACC and ACC).

	ACT	STA	ACC	ACH
ACT	1	0	1	0
STA	2	3	4	2
ACC	1	0	6	0
ACH	0	1	9	10
precision:	0.25	0.75	0.3	0.83
recall:	0.5	0.27	0.86	0.5
f-measure:	0.33	0.4	0.44	0.63
accuracy:	50%			

Table 10: Precision and recall results from experiment 5, with 4 categories

If we lump together ACH and ACC into a larger class of telic events (TEL), results (table 10) show a significative improvement.

	ACT	STA	TEL
precision:	0.25	0.75	0.78
recall:	0.5	0.27	0.93
f-measure:	0.33	0.4	0.85
accuracy:	72.5%		

Table 11: Precision and recall results from experiment 5, with 3 categories

5. Conclusions

Event type represents a key element of verb semantics. The interaction of various contextual factors in determining the event type expressed by a sentence makes event type identification a highly challenging task.
We have reported on two different models of event type classification, that have shown how both supervised and unsupervised approaches can account for the contribution of contextual features in identifying the sentence event type. Moreover, a human tagging experiment has shown how the task of event type identification is not trivial even for humans. This makes the performance achieved by the models reported above even more significant, if compared with human tagging accuracy.
Computational models can help us to shed new light on the real structure of event type classes as well as to gain a better understanding of context-driven semantic shifts. Stochastic algorithms appear to be a new and interesting way to model event types in a dynamic way, because they are able to grasp the complex interaction of contextual features by representing them as probabilistic cues to determine event types.

6. References

M. Baroni, S. Bernardini, F. Comastri, L. Piccioni, A. Volpi, G. Aston, and M. Mazzoleni. 2004. Introducing the la repubblica corpus: a large, annotated, tei(xml)-compliant corpus of newspaper italian. In F. Ferreira R. Costa R. Silva M. Lino, M. Xavier, editor, *Proceedings of LREC 2004*, pages 1771–1774, Lisbona. ELDA.

A. Berger, S. Della Pietra, and V. Della Pietra. 1996. A maximum entropy approach to natural language processing. *Computational Linguistics*, 22(1):39–71.

P. M. Bertinetto. 1986. *Tempo, Aspetto e Azione nel verbo italiano. Il sistema dell'indicativo*. Accademia della Crusca, Firenze.

B. Comrie. 1976. *Aspect: An Introduction to Verbal Aspect and Related Problems*. Cambridge University Press.

D. Dowty. 1979. *Word meaning and Montague Grammar. The semantics of verbs and times in generative semantics and in Montague's PTQ*. Reidel, Dordrecht.

S. T. Dumais. 1990. Enhancing performance in latent semantic indexing. TM-ARH 017527, Bellcore.

Z. S. Harris. 1968. *Mathematical Structures of Language*. Wiley, New York.

T. Kohonen. 1997. *Self-organizing Maps*. Springer, New York.

K. Lagus and A. Airola. 2005. Semantic clustering of verbs - analysis of morphosyntactic contexts using the som algorithm. In A. Lenci, S. Montemagni, and V. Pirrelli, editors, *Acquisition and Representation of Word Meaning: Theoretical and computational perspectives*, Linguistica Computazionale XXII-XXIII. IEPI, Pisa-Roma.

Y. Miyao. 2002. Maximum entropy estimation for feature forests. In *Proceedings of HLT2002*.

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, and R. Delmonte. 2003. Building the italian syntactic-semantic treebank. In A. Abeillé, editor, *Treebanks. Building and Using Parsed Corpora*, pages 189–210. Kluwer, Dortrecht.

A. Palmer, E. Ponvert, J. Baldridge, and C. Smith. 2007. A sequencing model for situation entity classification. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 896–903, Prague.

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

S. Rothstein. 2004. *Structuring Events - A Study in the Semantics of Lexical Aspect*. Blackwell Publishing.

R. Siegel and K. R. McKeown. 2000. Learning methods to combine linguistic indicators: improving aspectual classification and revealing linguistic insights. *Computational Linguistics*, 26(4):595–628.

Z. Vendler, 1967. *Linguistics in Philosophy*, chapter Verbs and Times, pages 97–121. Cornell University Press, Ithaca, NY.