Corpus-based Acquisition of German Event- and Object-Denoting Nouns

Stefan Gorzitze and Sebastian Padó
Institut für Computerlinguistik
Universität Heidelberg
gorzitze,pado@cl.uni-heidelberg.de

Abstract

This paper presents a simple distributional method for acquiring event-denoting and object-denoting nouns from corpora. Its core is a bootstrapping cycle that alternates between acquiring new instances and new features, using a simple log odds ratio for filtering. We acquire 3000 German nouns for each class with precisions of 93% (events) and 98% (objects), respectively.

1 Events and Objects

While the majority of nouns in English and related languages refer to objects, either physical (*book*) or abstract (*idea*), some nouns refer to events and states (*christening, happiness*).[1]

Being able to distinguish between these two classes is desirable for a number of reasons. From a model theoretic semantics point of view, event nouns and object should at the very least receive lexical entries of different types that mirror their different semantic behavior and associated information (event information vs. relational or qualia information). The distinction is also relevant for applications ranging from question answering (where entities and events usually correspond to different question types, cf. Hirschman and Gaizauskas 2000) to information extraction (where events are generally of primary importance, cf. Saurí et al. 2005) and to the modeling of reading times in psycholinguistics where the event/entity distinction often plays an important role (Traxler et al., 2002).

[1]For simplicity, we will refer to the first class as *object nouns* and the second class as *event nouns*. We acknowledge that the event/object dichotomy is an oversimplification; see the discussion in Peris et al. (2012).

There is a great deal of literature on nominalizations and their ambiguities, but relatively little work on the acquisition of event and object nouns (but see Eberle et al., 2009; Peris et al., 2012). Ontologies like WordNet (Miller et al., 1990) distinguish events and objects, but coverage remains a problem even for English. For other languages, such resources are typically much smaller. For languages like German, the additional problem of productive *compounding* arises: it is impossible to cover all noun compounds in an ontology.

This paper addresses this problem with a simple method for acquiring event and object nouns. Its core is a corpus-based bootstrapping cycle which exploits distributional differences between the two classes, alternating between instances and features. It is largely language-independent; our evaluation on German shows promising results.

2 A Corpus-based Acquisition Method

2.1 Distributional Features

Our intuition is that event nouns refer to entities which have a *temporal dimension*, while object nouns refer to entities that do not (Peris et al., 2012). This conceptual difference is mirrored in the usage of event and object nouns and can thus be picked up with distributional semantics (Turney and Pantel, 2010). Within distributional models, it has been observed (Peirsman et al., 2008) that "loose" contexts (i.e., large bag-of-word contexts) tend to pick up semantic *relatedness* while "tight" contexts (small bag-of-words contexts or syntactic contexts) modeling semantic *similarity* better. Since the event/object distinction belongs to the second type, and since all target words are nouns, we consider three types of *syntactic* contexts.

The first type covers direct modifiers of the target nouns, namely adjectives, which typically refer to properties of the nouns. Event nouns should therefore support adjectives that refer to temporal properties (*recent, frequent*) while object nouns occur with adjectives that refer to physical properties (*large, blind*). The second type covers occurrences of the target nouns as prominent arguments (subjects and objects) of verbs. Again, we expect that events occur preferably with verbs dealing with temporal or aspectual properties (occurring as the grammtical objects of transitive *begin, repeat, postpone*) while object verbs support a large variety of events (occurring as the grammatical objects of *drink, transport, love*). Finally, the third type covers occurrences of the target nouns embedded in larger NPs, such as *N of target*. As before, event nouns should occur in NPs with "temporal" heads such as *anniversary, period* while object nouns prefer nouns such as *collection, West*.

This approach comes with two main potential problems. The first one is the asymmetry between event and object nouns: event nouns should occur in restricted contexts, while object nouns support a wide variety of contexts. Furthermore, these contexts differ considerably among subgroups of object nouns (concrete vs. abstract objects). We will ignore this problem for the moment and assume a standard two-class classification process.

The second problem is the identification of predictive features. Clearly, using *all* verbs, adjectives, and nouns as features is infeasible. Furthermore, most of these features would be useless since they occur too infrequently, or because they can occur with both event and object nouns (e.g., *long* can refer to time length or physical dimensions). We require a method that can learn features directly from data and determine reliable ones.

2.2 A Bootstrapping Cycle

We approach the feature learning problem with the bootstrapping cycle shown in Figure 1. Bootstrapping has been applied to a variety of NLP tasks including lexical acquisition (Thelen and Riloff, 2002), question answering (Ravichandran and Hovy, 2002), and relation extraction (Pantel and Pennacchiotti, 2006). The idea is that knowledge about the nouns in a class can be used to acquire new features for the class, and vice versa.

Bootstrapping makes it possible to start out with a "seed" (in our case, a small set of either prototypical nouns or prototypical features) and acquire sets of nouns and features – in principle without further supervision. Crucial for the success of bootstrapping, however, is a sensible filtering mechanism, usually realized by way of a ranking function combined with selecting only the top n features and nouns, respectively. We follow Thelen and Riloff (2002) in taking advantage of a multi-class setup. Specifically, we score each feature f with respect to a class c using a simple log odds ratio: $\text{LOR}(f, c) = \log \frac{P(c|f)}{P(\neg c|f)}$ which measures how many times it is more likely that f indicates c than it does the other class.[2] In the inverse direction, we simply replace f by nouns n, employing the same formula to measure n's association with c: $\text{LOR}(n, c) = \log \frac{P(c|n)}{P(\neg c|n)}$.

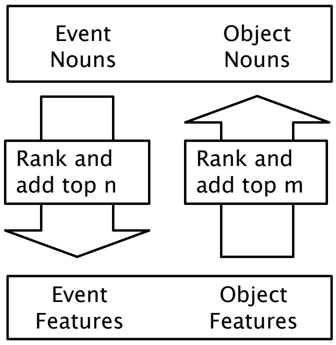

Figure 1: A bootstrapping cycle for learning event and object nouns as well as features for these classes

3 Evaluation

3.1 Setup

We tested our model on the sdewac corpus. sdewac is a subset of the deWac corpus (Baroni et al., 2008), a web-crawled German corpus containing about 1.9M documents from 11,000 different `.de` subdomains. sdewac was drawn from deWaC by removing duplicate and ungrammatical sentences, and parsing the remaining 900M tokens with a rule-based dependency parser (Schiehlen, 2003).

We seeded the bootstrapping cycle by manually specifying 100 noun lemmas for the event and object classes, respectively. To avoid that the cycle picks up features for specific domains rather than the event/object distinction, we included a wide range of nouns in both classes, based on the 26

[2]We use add-one smoothing.

	Events		Objects	
	IPrec	CPrec	IPrec	CPrec
Step 1	93.4	93.4	98.6	98.6
Step 2	93.5	93.5	98.0	98.3
Step 3	95.3	94.1	98.0	98.2
Step 4	93.6	94.0	95.2	97.5
Step 5	93.7	93.9	99.0	97.8
Step 6	93.1	93.8	99.1	98.0
Step 7	93.8	93.8	97.9	98.0
Step 8	92.1	93.6	98.6	98.1
Step 9	90.9	93.3	99.3	98.2
Step 10	90.0	92.9	99.0	98.3

Table 1: Precision of extracted event and object nouns: Values for individual batches (IPrec, 300 nouns each) and cumulative precision (CPrec)

WordNet "supersenses" / "lexicographer labels".

The cycle ran for 10 iterations over the whole corpus, with n (the number of features selected in each bootstrapping step) set to 150 and m (the number of lemmas) to 300. This resulted in 1500 features and 3000 lemmas for each class.

3.2 Method

Given that we do not have complete lists of event and object nouns, it is hard to compute recall. Similar to work in relation extraction (Pantel and Pennacchiotti, 2006), this paper focuses on a precision-based evaluation.

In order to gauge the difficulty of assigning nouns to the event and object categories, we performed a pre-experiment on a small dataset of 25 nouns which were annotated by 4 annotators each as either events, nouns, or ambiguous. Using Fleiss' κ (Fleiss, 1971), a measure of reliability appropriate for multiple annotators, we obtained an inter-annotator agreement of 0.76, which corresponds to substantial agreement. On the basis of this result, we decided to annotate the complete output of the method with single annotation, using the same annotation scheme as before. In the evaluation, we give full credit for each label of ambigous nouns, also for minority senses.

3.3 Results

Table 1 shows our results. The two columns for each class list the individual precision of the 300 nouns acquired in each step (IPrec) and the cumulative precision of all nouns up to this step (CPrec). The results are fairly high across the board. With figures around 98%, object nouns are easier to acquire than event nouns (in the low 90s), which is unsurprising since they form the majority class.[3] The precision of objects remains at a high level for all ten steps. For events, IPrec remains between 93% and 94% up until step 7. It then begins to drop, however. It appears that current settings work well to acquire a "core set" of some 2000 events but becomes more unreliable afterwards.

3.4 Analysis

Table 2 shows a random sample of acquired nouns for both classes, including occasional errors such as **Religionsausübung (religious observance)* identified as an object. The high accuracy of the objects is due at least partly to the large number of concrete nouns in the object class which are easy to categorize. In contrast, abstract nouns are relatively rare. The list also contains a substantial number of compounds, highlighting the benefits of distributional analysis for this class.

Table 3 shows a sample of acquired features, which bear out our assumptions (cf. Section 2.1) rather well. Among the verb features for events, we find a number of aspectual verbs (*dauern / take time*) as well as of "scheduling" verbs (*vorverlegen / move forward*). Object nouns occur in agent-like positions (as grammatical subjects) and as grammatical objects of causative verbs (*waschen / wash*). Some of the nominal features are nominalizations of verbs (*Ableistung / serving, e.g. a jail sentence*). These are complemented by temporal nouns for events (*Vorabend / previous evening*) and person and physical position nouns for objects (*Bürgermeister / mayor*). Finally, almost all adjective features for events refer to durations or to participants of the event (*amtsärztlich / by an officially appointed doctor*) while adjective features for objects mostly express physical properties (*rund / round*) or are adjectival passives ("Zustandspassive", *erworben / purchased*).

Finally, we sampled 800 correctly recognized nouns (400 events and 400 objects) and reclassified the nouns using three models that used just one feature type each. Table 4 evaluates them against the original 800-noun list. The noun model shows the worst results. The main culprit is a low

[3] In a random 100-word sample from the corpus, we found 56 object nouns, 32 event nouns, and 12 ambiguous nouns.

Events	Objects
Jahrestagung, Rundreise, Überprüfung, Exekution, Militärputsch, Auftauchen, Bergung, Pubertät, Freiheitsstrafe, Wiederkehr, Niederschlagung, Militäraktion, Wiedereröffnung, Auszählung, Praxissemester, Prophylaxe, Umfrage, Vorbereitungsphase, Feierstunde, Boom, Planungsphase, Währungsreform, Ratifizierung, Klausurtagung, 90er-Jahr, Machtkampf, Ersatzdienst, Tagung, Volksabstimmung, Ritt, Ultraschalluntersuchung	Antenne, Medaille, Nase, Mitmensch, Steuerzahler, Nadel, *Religionsausübung, Häuschen, Schlüssel, Wirtschaftsguts, Auge, Segel, Deckel, Nachbarstaat, Ministerin, Kapelle, Gefäß, Krankenkasse, Hase, Handschuh, Mitgliedsland, Tarifpartei, Konfliktpartei, Passagier, Beschwerdeführer, Linse, Schürze, Schwan, *Handlungsfähigkeit , Gebietskörperschaft, Flair, Fötus, 5tel, Ärztekammer , Elefant, Mehrheit, Gesundheitsamt, Eisenbahnlinie,

Table 2: Sample of acquired nouns

Verb features (events)	Verb features (objects)
anzetteln-OBJ, ableisten-OBJ, verstreichen-SUBJ, mitschneiden-OBJ, vertagen-OBJ, anberaumen-OBJ, jähren-SUBJ, verbüßen-OBJ, vorverlegen-OBJ, dauern-SUBJ	trinken-OBJ, erkranken-SUBJ, errichten-OBJ, investieren-SUBJ, engagieren-SUBJ, schütteln-OBJ, waschen-OBJ, erwerben-SUBJ, schöpfen-OBJ, spenden-OBJ, transportieren-OBJ
Noun features (events)	**Noun features (objects)**
Ableistung, Beendigung, Vorabend, Anmelder, Wirren, Schlußphase, Gräuel, Ausbruch, Veteran, Jahrestag, Vortag, Siegermacht, Versäumung, Verbüßung, Zurückverweisung, Ablegung, Ablauf	Seele, Osten, Beziehung, Ministerpräsident, Bürgermeister, Gründung, Erwerb, Auge, Wiederaufbau, Köpfen, Anstalt, Einwohner, Sohn, Wettbewerbsfähigkeit, Verkauf, Verabschiedung
Adjective features (events)	**Adjective features (objects)**
30jährig, erkennungsdienstlich, mündlich, medikamentös, anderweitig, monatelang, konzertant, amtsärztlich, ambulant, wochenlang, menschenunwürdig, physiotherapeutisch	mittelständisch, rund, behindert, rot, kreisfrei, ätherisch, tätig, erworben, gehandelt, hell, schwerbehindert, arm, beruflich, blau, interessiert, elektrisch, teilen, blind, gebildet, golden, ansässig

Table 3: Sample of acquired features

recall due to the low frequency of the "embedded NP" construction (cf. Section 2.1), but the precision is also imperfect: many nouns can embed events as well as objects (*Beziehung, Anmelder (relation, registrant)*). The verb model works substantially better. Notably, it has an almost perfect precision – there are verbs almost all subject (or objects, respectively) of which belong to one of the two classes. However, its recall is still fairly low. The best model is the adjective-based one, with the highest recall and an almost equal precision.

Open questions include the influence of target word frequency and domain. Given the space constraints, these must be left for future work.

	Recall	Precision	F_1
Adjective features	82.89	92.86	87.59
Verb features	61.66	95.46	74.92
Noun features	17.80	72.95	28.62
All features	100	100	100

Table 4: Results of feature ablation analysis (evaluation measures relative to full model)

4 Conclusion

We have presented a simple corpus-based method to learn event-denoting nouns and object-denoting nouns from corpora using a bootstrapping cycle to alternately acquire nouns and context features. Our method shows good results on German, but is essentially language-independent. It requires only a large parsed corpus and a seed set of nouns from both classes. Our feature ablation analysis indicates that full parsing may even be dispensable, as long as Adj-N-pairs can be identified reliably.

In the current paper, we have mostly ignored the issue of ambiguous nouns. In future work, we plan to apply our model to the disambiguation of nouns instances (rather than lemmas), which will involve considerably more sparsity.

Acknowledgments. This work was partially supported by the EC-funded project EXCITEMENT (FP7 ICT-287923).

References

Marco Baroni, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. 2008. The WaCky Wide Web : A Collection of Very Large Linguistically Processed Web-Crawled Corpora. *Language Resources and Evaluation*, 43(3):209–226.

Kurt Eberle, Gertrud Faaß, and Ulrich Heid. 2009. Corpus-based identification and disambiguation of reading indicators for german nominalizations. In *Proceedings of Proceedings of the 5th Corpus Linguistics Conference*, Liverpool, UK.

Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378–382.

Lynette Hirschman and Robert Gaizauskas. 2002. Natural language question answering: the view from here. *Natural Language Engineering*, 7(4):275–300.

George Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller. 1990. Introduction to WordNet: An On-Line Lexical Database. *International Journal of Lexicography*, 3(4):235–244.

Patrick Pantel and Marco Pennacchiotti. 2006. Espresso: Leveraging Generic Patterns for Automatically Harvesting Semantic Relations. In *Proceedings of COLING/ACL*, pages 113–120, Sydney, Austrial.

Yves Peirsman, Kris Heylen, and Dirk Geeraerts. 2008. Size matters: tight and loose context definitions in English word space models. In *Proceedings of the ESSLLI Workshop on Distributional Lexical Semantics*, pages 34–41.

Aina Peris, Mariona Taulé, and Horacio Rodríguez. 2012. Empirical methods for the study of denotation in nominalizations in spanish. *Computational Linguistics*. To appear.

Deepak Ravichandran and Eduard H Hovy. 2002. Learning surface text patterns for a Question Answering System. In *Proceedings of ACL*, Philadelphia, PA.

Roser Saurí, Robert Knippen, Marc Verhagen, and James Pustejovsky. 2005. Evita: a robust event recognizer for QA systems. In *Proceedings of HLT/EMNLP*, pages 700–707, Vancouver, BC.

Michael Schiehlen. 2003. A cascaded finite-state parser for German. In *Proceedings of EACL*, pages 163–166, Budapest, Hungary.

Michael Thelen and Ellen Riloff. 2002. A Bootstrapping Method for Learning Semantic Lexicons using Extraction Pattern Contexts. In *Proceedings of EMNLP*, pages 214–221, Philadelphia, PA.

Matthew J Traxler, Martin J Pickering, and Brian McElree. 2002. Coercion in sentence processing: evidence from eye-movements and self-paced reading. *Journal of Memory and Language*, 47(4):530–547.

Peter D Turney and Patrick Pantel. 2010. From Frequency to Meaning: Vector Space Models of Semantics. *Artificial Intelligence*, 37(1):141–188.