

A machine learning approach to relational noun mining in German

Berthold Crysmann

Arbeitsbereich Sprache und Kommunikation

Universität Bonn

crysmann@uni-bonn.de

Abstract

In this paper I argue in favour of a collocation extraction approach to the acquisition of relational nouns in German. We annotated frequency-based best lists of noun-preposition bigrams and subsequently trained different classifiers using (combinations of) association metrics, achieving a maximum F-measure of 69.7 on a support vector machine (Platt, 1998). Trading precision for recall, we could achieve over 90% recall for relational noun extraction, while still halving the annotation effort.

1 Mining relational nouns: almost a MWE extraction problem

A substantial minority of German nouns are characterised by having an internal argument structure that can be expressed as syntactic complements. A non-negligible number of relational nouns are deverbal, inheriting the semantic argument structure of the verbs they derive from. In contrast to verbs, however, complements of nouns are almost exclusively optional.

The identification of relational nouns is of great importance for a variety of content-oriented applications: first, precise HPSG parsing for German cannot really be achieved, if a high number of noun complements is systematically analysed as modifiers. Second, recent extension of Semantic Role Labeling to the argument structure of nouns (Meyers et al., 2004) increases the interest in lexicographic methods for the extraction of noun subcategorisation information. Third, relational nouns are also

a valuable resource for machine translation, separating the more semantic task of translating modifying prepositions from the more syntactic task of translating subcategorised for prepositions. Despite its relevance for accurate deep parsing, the German HPSG grammar developed at DFKI (Müller and Kasper, 2000; Crysmann, 2003; Crysmann, 2005) currently only includes 107 entries for proposition taking nouns, and lacks entries for PP-taking nouns entirely.

In terms of subcategorisation properties, relational nouns in German can be divided up into 3 classes:

- nouns taking genitival complements (e.g., *Beginn der Vorlesung* ‘beginning of the lecture’, *Zerstörung der Stadt* ‘destruction of the city’)
- nouns taking propositional complements, either a complementiser-introduced finite clause (*der Glaube, daß die Erde flach ist* ‘the belief that earth is flat’), or an infinitival clause (*die Hoffnung, im Lotto zu gewinnen* ‘the hope to win the lottery’), or both
- nouns taking PP complements

In this paper, I will be concerned with nouns taking prepositional complements, although the method described here can also be easily applied to the case of complementiser-introduced propositional complements.¹

¹In fact, I expect the task of mining relational nouns taking finite propositional complements to be far easier, owing to a reduced ambiguity of the still relatively local complementiser

The prepositions used with relational nouns all come from a small set of basic prepositions, mostly locative or directional.

A characteristic of these prepositions when used as a noun's complement, is that their choice becomes relatively fixed, a property shared with MWEs in general. Furthermore, choice of preposition is often arbitrary, sometimes differing between relational nouns and the verbs they derive from, e.g., *Interesse an* 'lit: interest at' vs. *interessieren für* 'lit: to interest for'. Owing to the lack of alternation, the preposition by itself does not compositionally contribute to sentence meaning, its only function being the encoding of a thematic property of the noun. Thus, in syntacto-semantic terms, we are again dealing with prototypical MWEs.

The fact that PP complements of nouns, like modifiers, are syntactically optional, together with the fact that their surface form is indistinguishable from adjunct PPs, makes the extraction task far from trivial. It is clear that grammar-based error mining techniques (van Noord, 2004; Cholakov et al., 2008) that have been highly successful in other areas of deep lexical acquisition (e.g., verb subcategorisation) cannot be applied here: first, given that an alternative analysis as a modifier is readily available in the grammar, missing entries for relational nouns will never incur any coverage problems. Furthermore, since PP modifiers are highly common we cannot expect a decrease in tree probability either.

Instead, I shall exploit the MWE-like properties of relational nouns, building on the expectation that the presence of a subcategorisation requirement towards a fixed, albeit optional, prepositional head should leave a trace in frequency distributions. Thus, building on previous work in MWE extraction, I shall pursue a data-driven approach that builds on a variety of association metrics combined in a probabilistic classifier. Despite the difference of the task,

daß. Although complement that-clauses in German can indeed be extraposed, corpus studies on relative clause extraposition (Uszkoreit et al., 1998) have shown that the great majority of extrapositions operates at extremely short surface distance, typically crossing the verb or verb particle in the right sentence bracket. Since locality conditions on complement clause extraposition are more strict than those for relative clause extraposition (Kiss, 2005; Crysman, to appear), I conjecture that the actual amount of non-locality found in corpora will be equally limited.

the approach suggested here shares some significant similarity to previous classifier-based approaches to MWE (Pecina, 2008).

2 Data

2.1 Data preparation

As primary data for relational noun extraction, I used the deWaC corpus (Baroni and Kilgariff, 2006), a 1.6 billion token corpus of German crawled from the web. The corpus is automatically tagged and lemmatised by TreeTagger (Schmid, 1995). From this corpus, I extracted all noun (NN) and preposition (APPR) unigrams and noun-preposition bigrams. Noun unigrams occurring less than ten times in the entire corpus were subsequently removed. In addition to the removal of hapaxes, I also filtered out any abbreviations.

Frequency counts were lemma-based, a decision that was motivated by the intended application, namely mining of relational noun entries for a lemma-based HPSG lexicon.

From the corpus, I extracted a best-list, based on bigram frequency, a well-established heuristical measure for collocational status (Krenn, 2000). Using a frequency based best list not only minimises initial annotation effort, but also ensures the quickest improvement of the target resource, the grammar's lexicon. Finally, the use of ranked best lists will also ensure that we will always have enough positive items in our training data.

2.2 Annotation

The ranked best list was subsequently annotated by two human annotators (A1,A2) with relatively little prior training in linguistics. In order to control for annotation errors, the same list was annotated a second time by a third year student of linguistics (A3).

In order to operationalise the argument/modifier annotators were asked to take related verbs into consideration, as well as to test (local and temporal) prepositions for paradigmatic interchangeability. Furthermore, since we are concerned with logical complements of nouns but not possessors, which can be added quite freely, annotators were advised to further distinguish whether a *von*-PP was only possible as a possessor or also as a noun complement.

An initial comparison of annotation decisions

showed an agreement of .82 between A1 and A3, and an agreement of .84 between A2 and A3. In a second round discrepancies between annotators were resolved, yielding a gold standard annotation of 4333 items, out of which 1179 (=27.2%) were classified as relational nouns.

3 Experiments

All experiments reported here were carried out using WEKA, a Java platform for data exploration and experimentation developed at the University of Waikato (Hall et al., 2009).

Since our task is to extract relational nouns and since we are dealing with a binary decision, performance measures given here report on relational nouns only. Thus, we do not provide figures for the classification of non-relational nouns or any uninformative (weighted) averages of the two.²

3.1 Learners

In a pre-study, we conducted experiments with a single feature set, but different classifiers in order to determine which ones performed best on our data set. Amongst the classifiers we tested were 2 Bayesian classifiers (Naive Bayes and Bayesian Nets), a Support Vector Machine, a Multilayer Perceptron classifier, as well as the entire set of decision tree classifiers offered by WEKA 3.6.4 (cf. the WEKA documentation for an exhaustive list of references). All test runs were performed with default settings. Unless otherwise indicated, all tests were carried out using 10-fold cross-validation.

Among these, decision tree classifiers perform quite well in general, with NBTree, a hybrid decision tree classifier using Naive Bayes classifiers at leaf nodes producing optimal results. Performance of the Naive Bayes classifier was suboptimal, with respect to both precision and recall. Overall performance of the Bayesian Net classifier (with a K2 learner) was competitive to average decision tree classifiers, delivering particularly good recall, but fell short of the best classifiers in terms of precision and F-measure.

²A base-line classifier that consistently choses the majority class (non-relational) and therefore does not detect a single relational noun, already achieves an F-measure for non-relational nouns of 84.3, and a weighted F-measure of 61.3%.

Thus, for further experimentation, we concentrated on the two best-performing classifiers, i.e., NBTree (Kohavi, 1996), which achieved the highest F-score and the second best precision, and SMO (Platt, 1998), a support vector machine, which produced the best precision value.

After experimentation regarding optimal feature selection (see next section), we re-ran our experiments with the modified feature set, in order to confirm that the classifiers we chose were still optimal. The results of these runs are presented in table 1.

	Prec.	Rec.	F-meas.
ADTree	68.3	61.1	64.5
BFTree	75.0	51.7	61.2
DecisionStump	52.5	80.2	63.5
FT	73.8	59.1	65.7
J48	72.9	58.4	64.8
J48graft	72.6	58.4	64.7
LADTree	70.5	57.5	63.3
LMT	74.9	59.8	66.5
NBTree	74.9	62.8	68.7
RandomForest	67.4	63.4	65.3
RandomTree	61.8	61.1	61.4
REPTree	74.5	61.2	67.2
Naive Bayes	70.5	53.9	61.1
Bayes Net	60.6	71.4	65.6
SMO	76.5	57.7	65.8
MultilayerPerceptron	67.5	64.5	65.9
Bagging (RepTree)	75.9	62.4	68.5
Voting (maj)	72.7	66.3	69.4
Voting (av)	71.3	68.4	69.8

Table 1: Performance of different classifiers

Finally, we did some sporadic test using a voting scheme incorporating 3 classifiers with high precision values (SMO, NBTree, Bagging(REPTree) (Breiman, 1996)), as well as two classifiers with high recall (BayesNet, recall-oriented SMO, see below). Using averaging, we managed to bring the F-measure up to 69.8, the highest value we measured in all our experiments.

3.2 Features

For NBTree, our best-performing classifier, we subsequently carried out a number of experiments to assess the influence and predictive power of individual association measures and to study their interactions.

Essentially, we make use of two basic types of features: string features, like the form of the preposition or the prefixes and suffixes of the noun, and association measures. As for the latter, we drew on the set of measures successfully used in previous studies on collocation extraction:

Mutual information (MI) An information theoretic measure proposed by (Church and Hanks, 1990) which measures the joint probability of the bigram in relation to the product of the marginal probabilities, i.e., the expected probability.

$$MI = \frac{p(\textit{noun}, \textit{prep})}{p(\textit{noun}) * p(\textit{prep})}$$

MI² A squared variant of mutual information, previously suggested by (Daille, 1994). Essentially, the idea behind squaring the joint probability is to counter the negative effect of extremely low marginal probabilities yielding high MI scores.

$$MI^2 = \frac{(p(\textit{noun}, \textit{prep}))^2}{p(\textit{noun}) * p(\textit{prep})}$$

Likelihood ratios A measure suggested by (Dunning, 1993) that indicates how much more likely the cooccurrence is than mere coincidence.

$$LR = \log L(p_1, k_1, n_1) + \log L(p_2, k_2, n_2) - \log L(p, k_1, n_1) - \log L(p, k_2, n_2)$$

where

$$\log L(p, n, k) = k \log p + (n - k) \log(1 - p)$$

and

$$p_1 = \frac{k_1}{n_1}, p_2 = \frac{k_2}{n_2}, p = \frac{k_1 + k_2}{n_1 + n_2}$$

t-score The score of Fisher's t-test. Although the underlying assumption regarding normal distribution is incorrect (Church and Mercer, 1993), the score has nevertheless been used with repeated success in collocation extraction tasks (Krenn, 2000; Krenn and Evert, 2001; Evert and Krenn, 2001).

$$t\textit{score} = \frac{p(\textit{noun}, \textit{prep}) - (p(\textit{noun}) * p(\textit{prep}))}{\sqrt{\frac{\sigma^2}{N}}}$$

As suggested by (Manning and Schütze, 1999) we use p as an approximation of σ^2 .

Association Strength (Smadja, 1993)

A factor indicating how many times the standard deviation a bigram frequency differs from the average.

$$Strength = \frac{freq_i - \bar{f}}{\sigma}$$

Best Indicates whether a bigram is the most frequent one for the given noun or not.

Best-Ratio A relative version of the previous feature indicating the frequency ratio between the current noun-preposition bigram and the best bigram for the given noun.

In addition to the form of the preposition, we included information about the noun's suffixes or prefixes:

Noun suffix We included common string suffixes that may be clues as to the relational nature of the noun, as, e.g., the common derivational suffixes *-ion*, *-schaft*, *-heit*, *-keit* as well as the endings *-en*, which are found inter alia with nominalised infinitives, and *-er*, which are found, inter alia with agentive nominals. All other suffixes were mapped to the NONE class.

Noun prefix Included were prefixes that commonly appear as verb prefixes. Again, this was used as a shortcut for true lexical relatedness.

As illustrated by the diagrams in Figure 1, the aforementioned association measures align differently with the class of relational nouns (in black):

The visually discernible difference in alignment between association metrics and relational nouns was also confirmed by testing single-feature classifiers: as detailed in Table 2, *MI*, *MI²*, and *t-score* all capable to successfully identify relational nouns by themselves, whereas *best*, *best-ratio* and *strength*

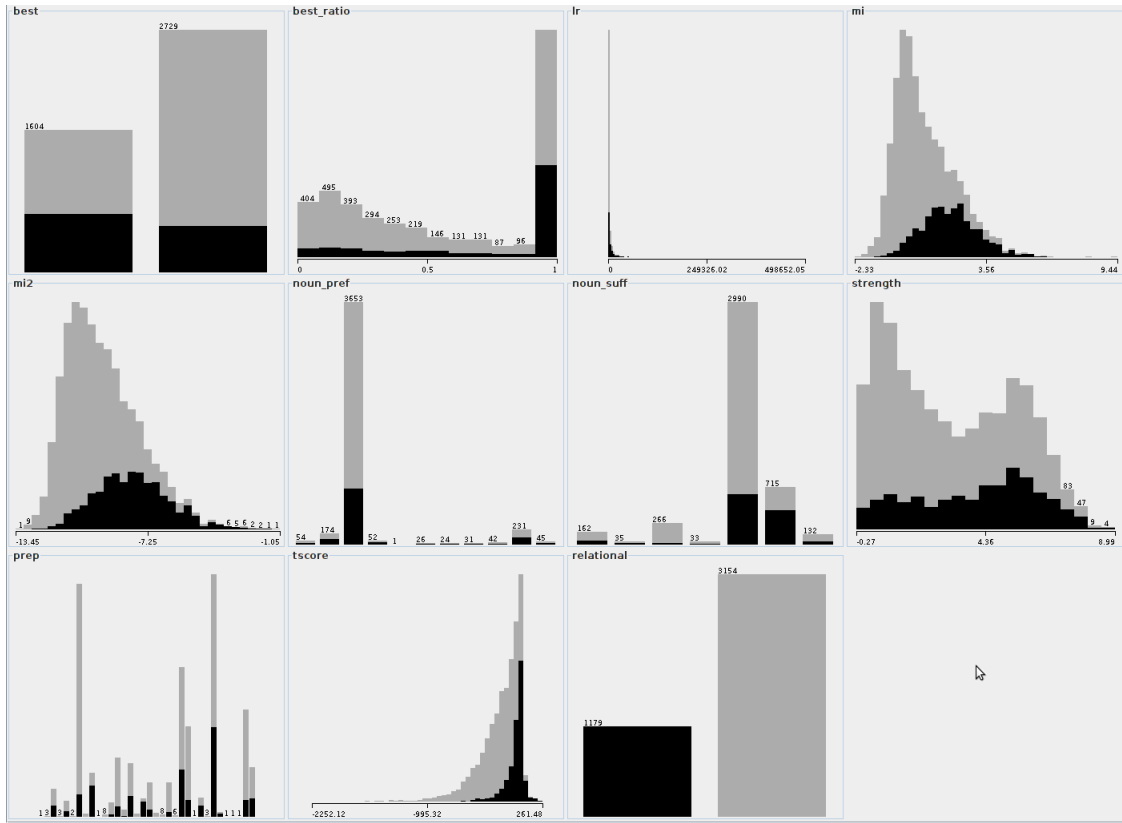


Figure 1: Distribution of relational and non-relational nouns across features (created with WEKA 3.6.4)

are entirely unable to partition the data appropriately. *LR* assumes an intermediate position, suffering mainly from recall problems.

	Prec.	Rec.	F-meas.
MI	65.2	45.2	53.4
MI2	62.2	50.7	55.9
LR	60	23.5	33.8
T-score	66.4	42	51.5
Strength	0	0	0
Best	0	0	0
Best-Ratio	0	0	0

Table 2: Classification by a single association metric

The second experiment regarding features differs from the first by the addition of form features:

Two things are worth noting here: first, the values achieved by *MI* and *T-score* now come very close to the values obtained with much more elaborate feature sets, confirming previous results on the usefulness of these metrics. Second, all association measures now display reasonable performance. Both

	Prec.	Rec.	F-meas.
MI	74.2	61.2	67.1
MI2	72.5	56.4	63.5
LR	73.1	54.4	62.4
T-score	74.9	60.6	67
Strength	72.5	52.4	60.9
Best	69.7	48.7	57.3
Best-Ratio	72.1	53.4	61.3

Table 3: Classification by a single association metric + form features (*preposition*, *noun prefix*, *noun suffix*)

these effects can be traced to a by-category sampling introduced by the form features. The most clear-cut case is probably the *best* feature: as shown in Figure 1, there is a clear increase in relational nouns in the TRUE category of the Boolean *best* feature, yet, they still do not represent a majority. Thus, a classifier with a balanced cost function will always prefer the majority vote. However, for particular noun classes (and prepositions for that matter) majorities can be tipped.

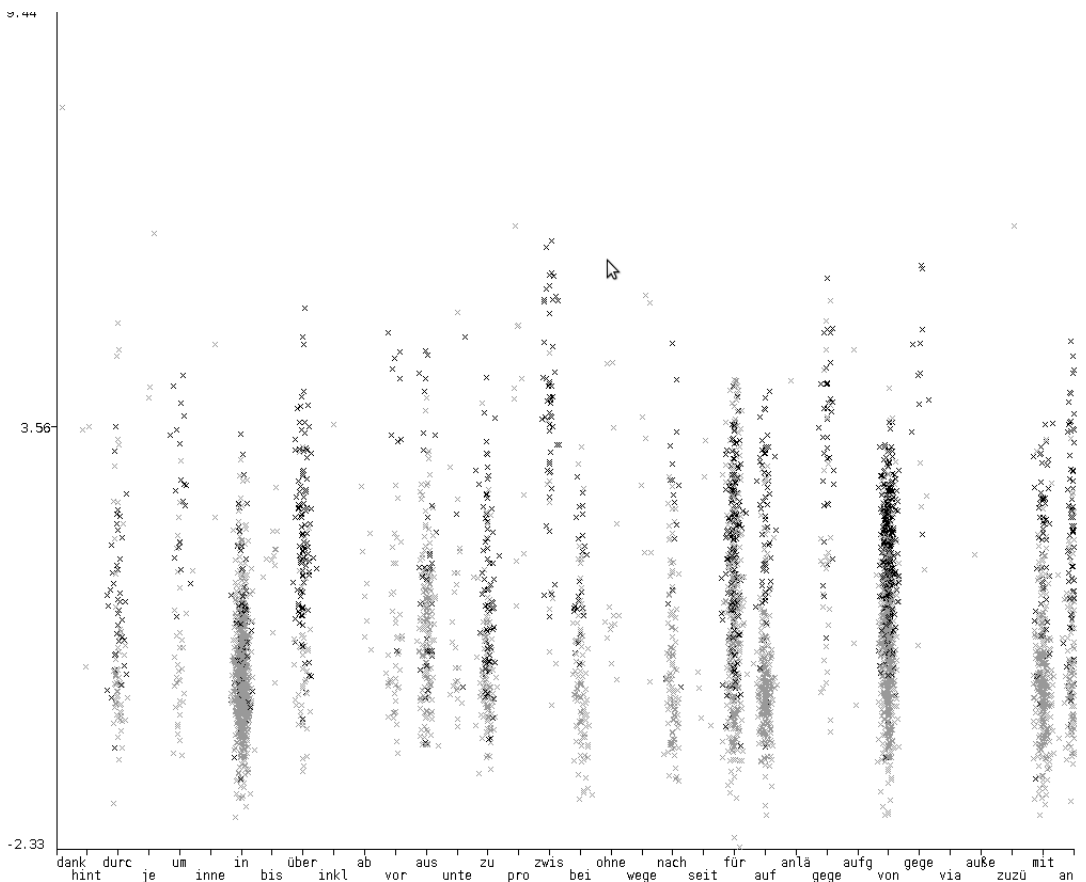


Figure 2: MI-values of relational nouns relative to preposition

As depicted by the preposition-specific plot of MI values in Figure 2, some prepositions have a clear bias for their use with relational nouns (e.g., *von* ‘of’) or against it (e.g., *ab* ‘from’), while others appear non-committal (e.g., *für* ‘for’). Similar observations can be made for noun suffixes and prefixes.

The next set of experiments were targetted at optimisation. Assuming that the candidate sets selected by different metrics will not stand in a subset relation I explored which combination of metrics yielded the best results. To do this, I started out with a full set of features and compared this to the results obtained with one feature left out. In a second and third step of iteration, I tested whether simultaneously leaving out some features for which we observed some gain would produce an even more optimised classifier.

Table 4 presents the result of the first step. Here, two outcomes are of particular interest: deleting information about the noun suffix is detrimental,

	Prec.	Rec.	F-meas.
All	74.4	61.2	67.2
–T-score	75.3	62.4	68.3
–MI	72.8	62.3	67.1
–MI ²	75.1	61.6	67.7
–LR	74.1	60.1	66.3
–Strength	73.4	62	67.2
–Best	73.7	60.7	66.6
–Best-Ratio	74.2	61.8	67.4
–Prep	74.7	61.1	67.2
–Noun-Prefix	74.7	61.1	67.2
–Noun-Suffix	71.3	55.3	62.3

Table 4: Effects of leaving one feature out

whereas ignoring the t-score value appears to be beneficial to overall performance.

In a second (and third) iteration, I tested whether any additional feature deletion apart from *t-score* would give rise to any further improvements.

–t-score	Prec.	Rec.	F-meas.
	75.3	62.4	68.3
–MI	74.4	57.6	64.9
–LR	74.8	61.3	67.4
–MI2	74.1	61.7	67.4
–Strength	75.1	62.8	68.4
–Best	74.1	61.5	67.2
–Best-Ratio	75.4	62.6	68.4
–Best-Ratio –Strength	74.9	63.4	68.7

Table 5: Effects of leaving two or more features out

In fact, removal of the *Strength* feature provided good results, whether taken out individually or in combination, which may be due to this feature’s inherently poor statistical properties (cf. Figure 1). Ignoring *best-ratio* was also beneficial, probably due to the fact that most of its beneficial properties are already covered by the *best* feature and that non-best noun-preposition combinations hardly ever give rise to positive hits.

As a matter of fact, simultaneous removal of *best-ratio* and *strength*, in addition to the removal of *t-score* of course, yielded best overall results. As a consequence, all remaining test runs were based on this feature set. In separate test runs with the SMO classifier, I finally confirmed that the optimality of this feature set was not just an artifact of the classifier, but that it generalises to SVMs as well.

3.3 Trade-offs

Since our main aim in relational noun mining is the improvement of the accuracy of our grammar’s lexicon, and since the quickest improvement are expected for highly frequent noun-preposition bigrams, I tested whether I could bring the recall of our classifiers up, at the expense of moderate losses in precision. For this evaluation, I used again our best-performing classifier (NBTree), as well as SMO, which had the highest head-room in terms of precision, while already providing satisfactory recall. To this end, I manipulated the classifier’s cost matrix during training and testing, gradually increasing the costs for false negatives compared to false positives.

The results of this evaluation are given in Figure 3. First, we obtained a new optimal f-measure for the SMO classifier: at a cost factor of 2.1 for false negatives, the f-measure peaks at 69.7, with a recall

of 75.1% and precision still acceptable (65.1%). At this level, we still save more than two thirds of the annotation effort.

By way of penalising false negatives 6 times more than false positives, the support vector machine was able to detect over 90% of all relational nouns, at a precision of 50%. At these levels, we can still save more than half of the entire annotation effort.

Going further down the Zipf distribution, we expect the savings in terms of annotation effort to go further up, since our bigram frequency ranking ensures that relational nouns are overrepresented at the top of the list, a rate that will gradually go down.

Finally, including false positives in the data to be annotated will also ensure that we always have enough positive and negative training data for learning a classifier on an extended data set.

3.4 Outlook

Although results are already useful at this point, I hope to further improve precision and recall rates by means of additional features. Evaluating the NBTree classifier on the training data, we observe an F-measure of only 74.7%, which suggests that the current set of features models the training data still quite imperfectly. Thus, one needs to incorporate further independent evidence in order to predict relation nouns more reliably. Owing to the semantic nature of the relational vs. non-relational distinction one type of additional evidence could come from multilingual resources: as a first step, I envisage incorporating the classification of nouns in the English Resource Grammar (ERG; (Copestake and Flickinger, 2000)) as prior information regarding relational status. In a second step I shall explore whether one can exploit information from parallel corpora, using in particular item-specific divergence of preposition choice to detect whether we are dealing with a contentful or rather a functional preposition.³ The intuition behind using cross-linguistic evidence to try and boost the performance of the learner is based on the observation that predicate argument structure in closely related languages such as English and German tends to be highly similar, with differences mostly located in syntactic proper-

³I expect that arbitrary divergence in the choice of preposition provides an indicator of grammaticalisation.

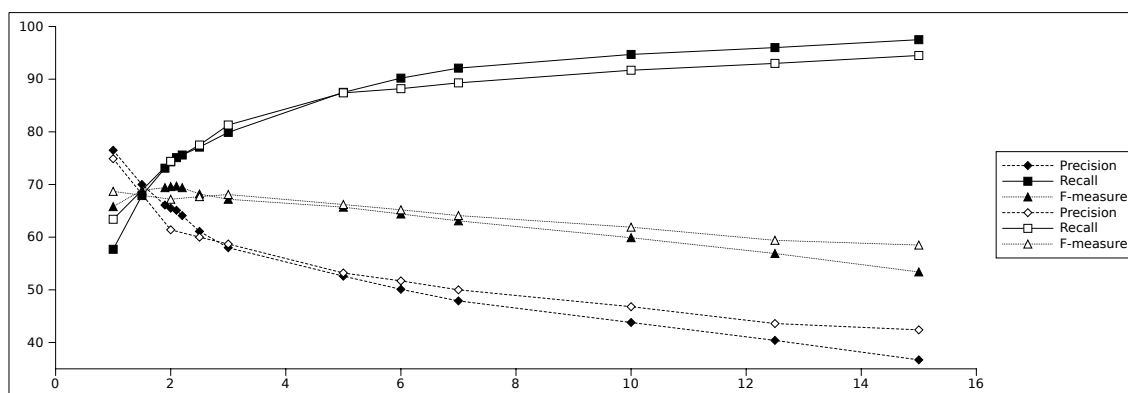


Figure 3: Effect of trading precision for recall (NBTrees: white; SMO: black)

ties such as selection for case or choice of preposition. As a consequence, I do not expect to be able to predict the actual form of the German preposition, but rather gain additional evidence as to whether a given noun has some relational use at all or not.

The second type of information that I plan to use more systematically in the future is morphological and lexical relatedness which is only approximated at present by the noun suffix and noun prefix features which hint at the derived (deverbal) nature of the noun under discussion. In addition to these brute-force features, I plan to incorporate the HPSG grammar’s verb subcategorisation lexicon, pairing nouns and verbs by means of minimum edit distance.⁴ In essence, we hope to provide a more general approach to lexical relatedness between relational nouns and the non-unary verbal predicates they derive from: in the current feature set, this was only suboptimally approximated by the use of noun suffix and prefix features, resulting in most nouns being mapped to the unproductive class NONE.⁵

Finally, I plan to apply the current approach to the extraction of nouns taking propositional complements. Given the comparative ease of that task compared to the extraction of PP-taking nouns, I shall investigate whether we can exploit the fact that many

⁴Being aware of the fact that lexical derivation may give rise to arbitrary changes in syntactic subcategorisation, I minimally expect to gather evidence regarding the arity of the derived noun predicate. To what extent actual selectional properties as to the shape of the functional preposition are maintained by derivational processes remains a matter of empirical research.

⁵The inclusion of noun prefixes, which are actually verb prefixes, is inherently limited to mimick lexical relatedness to prefix verbs.

relational nouns taking propositional complements (e.g., *der Glaube, daß* ... ‘the belief that’) also take PP-complements (*der Glaube an* ‘the belief in’) in order to further improve our present classifier. In a similar vein, I shall experiment whether it is possible to extrapolate from relational nouns taking *von*-PPs to genitive complements.

4 Conclusion

In this paper I have suggested to treat the task of mining relational nouns in German as a MWE extraction problem. Based on the first 4333 hand-annotated items of a best-list ranked by bigram frequencies, several classifiers have been trained in order to determine which learner and which (combination of) association measures performed best for the task.

Testing different classifiers and different metrics, we found that optimal results were obtained using a support vector machine (Platt, 1998), including Mutual Information (MI), its squared variant (MI^2), and Likelihood Ratios (LR) as association measures, together with information about the identity of the preposition and the noun’s prefix and suffix. The second best classifier, a hybrid decision tree with Naive Bayes classifiers at the leaves produced highly competitive results. T-scores, while being a good predictor on its own, however, led to a slight decrease in performance, when a full feature set was used. Likewise, performance suffered when Association Strength (Smadja, 1993) was included. Overall performance of the best individual classifier figured at an F-score of 69.7.

References

- Marco Baroni and Adam Kilgarriff. 2006. Large linguistically-processed web corpora for multiple languages. In *Proceedings of EACL 2006*.
- Leo Breiman. 1996. Bagging predictors. *Machine Learning*, 24(2):123–140.
- Kostadin Cholakov, Valia Kordoni, and Yi Zhang. 2008. Towards domain-independent deep linguistic processing: Ensuring portability and re-usability of lexicalised grammars. In *Coling 2008: Proceedings of the workshop on Grammar Engineering Across Frameworks*, pages 57–64, Manchester, England, August. Coling 2008 Organizing Committee.
- Kenneth Church and Patrick Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1):22–29.
- Kenneth Church and Robert Mercer. 1993. Introduction to the special issue on computational linguistics using large corpora. *Computational Linguistics*, 19:1–24.
- Ann Copestake and Dan Flickinger. 2000. An open-source grammar development environment and broad-coverage English grammar using HPSG. In *Proceedings of the Second conference on Language Resources and Evaluation (LREC-2000)*, Athens.
- Berthold Crysmann. 2003. On the efficient implementation of German verb placement in HPSG. In *Proceedings of RANLP 2003*, pages 112–116, Borovets, Bulgaria.
- Berthold Crysmann. 2005. Relative clause extraposition in German: An efficient and portable implementation. *Research on Language and Computation*, 3(1):61–82.
- Berthold Crysmann. to appear. On the locality of complement clause and relative clause extraposition. In Gert Webelhuth, Manfred Sailer, and Heike Walker, editors, *Rightward Movement in a Comparative Perspective*. John Benjamins, Amsterdam.
- Béatrice Daille. 1994. *Approche mixte pour l'extraction automatique de terminologie : statistique lexicale et filtres linguistiques*. Ph.D. thesis, Université Paris 7.
- Ted Dunning. 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19:61–74.
- Stefan Evert and Brigitte Krenn. 2001. Methods for the qualitative evaluation of lexical association measures. In *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics, Toulouse, France*, pages 188–195.
- Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. 2009. The WEKA data mining software: an update. *SIGKDD Explorations*, 11(1):10–18.
- Tibor Kiss. 2005. Semantic constraints on relative clause extraposition. *Natural Language and Linguistic Theory*, 23:281–334.
- Ron Kohavi. 1996. Scaling up the accuracy of naive-bayes classifiers: A decision-tree hybrid. In *Second International Conference on Knowledge Discovery and Data Mining*, pages 202–207.
- Brigitte Krenn and Stefan Evert. 2001. Can we do better than frequency? a case study on extracting PP-verb collocations. In *Proceedings of the ACL Workshop on Collocations, Toulouse, France*, pages 39–46.
- Brigitte Krenn. 2000. *The Usual Suspects: Data-oriented Models for the Identification and Representation of Lexical Collocations*. Ph.D. thesis, Universität des Saarlandes.
- Christopher Manning and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. MIT Press.
- A. Meyers, R. Reeves, C. Macleod, R. Szekeley, V. Zielinska, B. Young, and R. Grishman. 2004. The nombank project: An interim report. In A. Meyers, editor, *HLT-NAACL 2004 Workshop: Frontiers in Corpus Annotation*, pages 24–31, Boston, Massachusetts, USA, May 2 - May 7. Association for Computational Linguistics.
- Stefan Müller and Walter Kasper. 2000. HPSG analysis of German. In Wolfgang Wahlster, editor, *Verbmobil: Foundations of Speech-to-Speech Translation*, pages 238–253. Springer, Berlin.
- Pavel Pecina. 2008. A machine learning approach to multiword expression extraction. In *Proceedings of the LREC Workshop Towards a Shared Task for Multiword Expressions (MWE 2008)*, pages 54–61.
- J. Platt. 1998. Fast training of support vector machines using sequential minimal optimization. In B. Schoelkopf, C. Burges, and A. Smola, editors, *Advances in Kernel Methods - Support Vector Learning*. MIT Press.
- Helmut Schmid. 1995. Improvements in part-of-speech tagging with an application to German. In *Proceedings of the ACL SIGDAT-Workshop*, March.
- Frank Smadja. 1993. Retrieving collocations from text: Xtract. *Computational Linguistics*, 19(1):143–177.
- Hans Uszkoreit, Thorsten Brants, Denys Duchier, Brigitte Krenn, Lars Konieczny, Stephan Oepen, and Wojciech Skut. 1998. Studien zur performanzorientierten Linguistik. Aspekte der Relativsatzextraposition im Deutschen. *Kognitionswissenschaft*, 7:129–133.
- Gertjan van Noord. 2004. Error mining for wide coverage grammar engineering. In *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL'04), Barcelona, Spain*, pages 446–453.