Using a Probabilistic Class-Based Lexicon for Lexical Ambiguity Resolution

Detlef Prescher and **Stefan Riezler** and **Mats Rooth**
Institut für Maschinelle Sprachverarbeitung
Universität Stuttgart, Germany

Abstract

This paper presents the use of probabilistic class-based lexica for disambiguation in target-word selection. Our method employs minimal but precise contextual information for disambiguation. That is, only information provided by the target-verb, enriched by the condensed information of a probabilistic class-based lexicon, is used. Induction of classes and fine-tuning to verbal arguments is done in an unsupervised manner by EM-based clustering techniques. The method shows promising results in an evaluation on real-world translations.

1 Introduction

Disambiguation of lexical ambiguities in naturally occuring free text is considered a hard task for computational linguistics. For instance, word sense disambiguation is concerned with the problem of assigning sense labels to occurrences of an ambiguous word. Resolving such ambiguities is useful in constraining semantic interpretation. A related task is target-word disambiguation in machine translation. Here a decision has to be made which of a set of alternative target-language words is the most appropriate translation of a source-language word. A solution to this disambiguation problem is directly applicable in a machine translation system which is able to propose the translation alternatives. A further problem is the resolution of attachment ambiguities in syntactic parsing. Here the decision of verb versus argument attachment of noun phrases, or the choice for verb phrase versus noun phrase attachment of prepositional phrases can build upon a resolution of the related lexical ambiguities.

Statistical approaches have been applied successfully to these problems. The great advantage of statistical methods over symbolic-linguistic methods has been deemed to be their effective exploitation of minimal linguistic knowledge. However, the best performing statistical approaches to lexical ambiguity resolution themselves rely on complex information sources such as "lemmas, inflected forms, parts of speech and arbitrary word classes [...] local and distant collocations, trigram sequences, and predicate argument association" (Yarowsky (1995), p. 190) or large context-windows up to 1000 neighboring words (Schütze, 1992). Unfortunately, in many applications such information is not readily available. For instance, in incremental machine translation, it may be desirable to decide for the most probable translation of the arguments of a verb with only the translation of the verb as information source but no large window of surrounding translations available. In parsing, the attachment of a nominal head may have to be resolved with only information about the semantic roles of the verb but no other predicate argument associations at hand.

The aim of this paper is to use only minimal, but yet precise information for lexical ambiguity resolution. We will show that good results are obtainable by employing a simple and natural look-up in a probabilistic class-labeled lexicon for disambiguation. The lexicon provides a probability distribution on semantic selection-classes labeling the slots of verbal subcategorization frames. Induction of distributions on frames and class-labels is accomplished in an unsupervised manner by applying the EM algorithm. Disambiguation then is done by a simple look-up in the probabilistic lexicon. We restrict our attention to a definition of senses as alternative translations of source-words. Our approach provides a very natural solution for such a target-language disambiguation task—look for the most frequent target-noun whose semantics fits best with the

Class 19 PROB 0.0235		0.0250 room	0.0211 area	0.0125 world	0.0111 meeting	0.0106 range	0.0096 school	0.0085 service	0.0084 building	0.0081 road	0.0079 street	0.0076 market	0.0068 course	0.0067 doctor	0.0061 place	0.0059 part	0.0056 mind	0.0053 class	0.0050 scene	0.0049 path	0.0049 group	0.0048 work	0.0048 gray	0.0047 programme	0.0047 line	0.0046 life	0.0046 garden	0.0045 body	0.0044 miles	0.0043 system	0.0043 period
0.0629	enter.aso:o	•	•	•		•	•	•	•	•	•	•	•		•	•	•	•	•	•	•	•	•	•	•	•		•	•	•	•
0.0386	cover.aso:o		•	•		•	•		•	•	•	•	•	•	•	•		•		•	•	•	•	•	•			•		•	•
0.0321	call.aso:o	•		•		•		•		•	•				•		•	•		•			•	•	•	•		•		•	•
0.0236	include.aso:o	•	•	•		•	•	•		•			•		•		•	•		•	•	•		•	•			•	•	•	•
0.0226	run.aso:o	•				•							•		•					•	•	•			•	•		•			
0.0214	attend.aso:o				•		•		•						•		•	•			•			•							
0.0173	cross.aso:o	•	•		•		•		•	•	•		•		•					•			•		•			•			
0.0136	dominate.aso:o	•	•	•		•		•	•	•			•							•			•			•		•		•	•
0.0132	have.aso:s	•			•		•		•			•				•				•		•	•		•			•			•
0.0126	attract.aso:s		•		•		•		•				•							•	•		•		•			•			•
0.0124	occupy.aso:o	•	•	•		•				•			•							•			•			•		•		•	•
0.0115	include.aso:s	•	•	•		•			•							•				•	•		•		•			•		•	•
0.0113	contain.aso:s	•	•	•		•			•						•					•	•		•		•			•		•	•
0.0108	become.as:s	•							•				•							•			•		•			•		•	•
0.0099	form.aso:o		•		•			•					•							•			•		•			•			•
0.0086	collapse.as:s		•						•	•	•		•							•			•		•			•		•	•
0.0085	create.aso:o		•	•		•		•					•							•			•		•			•		•	•
0.0082	provide.aso:s	•				•		•							•					•			•		•			•		•	•
0.0082	organize.aso:o		•	•	•		•						•							•			•		•			•			•
0.0082	offer.aso:s	•	•	•		•		•		•	•	•	•	•		•				•			•		•			•			•

Figure 1: Class 19: "locative action". At the top are listed the 20 most probable nouns in the $p_{LC}(n|19)$ distribution and their probabilities, and at left are the 30 most probable verbs in the $p_{LC}(v|19)$ distribution. 19 is the class index. Those verb-noun pairs which were seen in the training data appear with a dot in the class matrix. Verbs with suffix $.as:s$ indicate the subject slot of an active intransitive. Similarily $.aso:s$ denotes the subject slot of an active transitive, and $.aso:o$ denotes the object slot of an active transitive.

semantics required by the target-verb. We evaluated this simple method on a large number of real-world translations and got results comparable to related approaches such as that of Dagan and Itai (1994) where much more selectional information is used.

2 Lexicon Induction via EM-Based Clustering

2.1 EM-Based Clustering

For clustering, we used the method described in Rooth et al. (1999). There classes are derived from distributional data—a sample of pairs of verbs and nouns, gathered by parsing an unannotated corpus and extracting the fillers of grammatical relations. The semantically smoothed probability of a pair (v, n) is calculated in a latent class (LC) model as $p_{LC}(v, n) = \sum_{c \in C} p_{LC}(c, v, n)$. The joint distribution is defined by $p_{LC}(c, v, n) = p_{LC}(c)p_{LC}(v|c)p_{LC}(n|c)$. By construction, conditioning of v and n on each other is solely made through the classes c. The parameters $p_{LC}(c)$, $p_{LC}(v|c)$, $p_{LC}(n|c)$ are estimated by a particularily simple version of the EM algorithm for context-free models.

Input to our clustering algorithm was a training corpus of 1,178,698 tokens (608,850 types) of verb-noun pairs participating in the grammatical relations of intransitive and transitive verbs and their subject- and object-fillers. Fig. 1 shows an induced class from a model with 35 classes. Induced classes often have a basis in lexical semantics; class 19 can be interpreted as *locative*, involving location nouns "room", "area", and "world" and verbs as "enter" and "cross".

2.2 Probabilistic Labeling with Latent Classes using EM-estimation

To induce latent classes for the object slot of a fixed transitive verb v, another statistical inference step was performed. Given a latent class model $p_{LC}(\cdot)$ for verb-noun pairs, and a sample $n_1, \ldots, n_M$ of objects for a fixed transitive verb, we calculate the probability of an arbitrary object noun $n \in N$ by $p(n) = \sum_{c \in C} p(c, n) = \sum_{c \in C} p(c)p_{LC}(n|c)$. This fine-tuning of the class parameters $p(c)$ to the sample of objects for a fixed verb is formalized again as a simple instance of the EM algorithm. In an experiment with English data, we used a clustering model with 35 classes. From the maximum probabil-

ity parses derived for the British National Corpus with the head-lexicalized parser of Carroll and Rooth (1998), we extracted frequency tables for transitive verb-noun pairs. These tables were used to induce a small class-labeled lexicon (336 verbs).

cross.aso:o 19	0.692	mobilize.aso:o 6	0.386
mind	74.2	force	2.00
road	30.3	people	1.95
line	28.1	army	1.46
bridge	27.5	sector	0.90
room	20.5	society	0.90
border	17.8	worker	0.90
boundary	16.2	member	0.88
river	14.6	company	0.86
street	11.5	majority	0.85
atlantic	9.9	party	0.80

Figure 2: Estimated frequencies of the objects of the transitive verbs *cross* and *mobilize*

Fig. 2 shows the topmost parts of the lexical entries for the transitive verbs *cross* and *mobilize*. Class 19 is the most probable class-label for the object-slot of *cross* (probability 0.692); the objects of *mobilize* belong with probability 0.386 to class 16, which is the most probable class for this slot. Fig. 2 shows for each verb the ten nouns n with highest estimated frequencies $f_c(n) = f(n)p(c|n)$, where $f(n)$ is the frequency of n in the sample $n_1, \ldots, n_M$. For example, the frequency of seeing *mind* as object of *cross* is estimated as 74.2 times, and the most frequent object of *mobilize* is estimated to be *force*.

3 Disambiguation with Probabilistic Cluster-Based Lexicons

In the following, we will describe the simple and natural lexicon look-up mechanism which is employed in our disambiguation approach. Consider Fig. 3 which shows two bilingual sentences taken from our evaluation corpus (see Sect. 4). The source-words and their corresponding target-words are highlighted in **bold face**. The correct translation of the source-noun (e.g. *Grenze*) as determined by the actual translators is replaced by the set of alternative translations (e.g. { border, frontier, boundary, limit, periphery, edge }) as proposed by the word-to-word dictionary of Fig. 5 (see Sect. 4).

The problem to be solved is to find a correct translation of the source-word using only minimal contextual information. In our approach,

(ID 160867) *Es gibt einige alte Passvorschriften, die besagen, dass man einen Pass haben muss, wenn man die* **Grenze** *überschreitet.* There are some old provisions regarding passports which state that people crossing the {**border**/ **frontier**/ **boundary**/ **limit**/ **periphery**/ **edge**} should have their passport on them.

(ID 201946) *Es gibt schliesslich keine Lösung ohne die Mobilisierung der bürgerlichen* **Gesellschaft** *und die Solidarität der Demokraten in der ganzen Welt.* There can be no solution, finally, unless civilian {**company**/ **society**/ **companionship**/ **party**/ **associate**} is mobilized and solidarity demonstrated by democrats throughout the world.

Figure 3: Examples for target-word ambiguities

the decision between alternative target-nouns is done by using only information provided by the governing target-verb. The key idea is to back up this minimal information with the condensed and precise information of a probabilistic class-based lexicon. The criterion for choosing an alternative target-noun is thus the best fit of the lexical and semantic information of the target-noun to the semantics of the argument-slot of the target-verb. This criterion is checked by a simple lexicon look-up where the target-noun with highest estimated class-based frequency is determined. Formally, choose the target-noun $\hat{n}$ (and a class $\hat{c}$) such that

$$f_{\hat{c}}(\hat{n}) = \max_{n \in N, c \in C} f_c(n)$$

where $f_c(n) = f(n)p(c|n)$ is the estimated frequency of n in the sample of objects of a fixed target-verb, $p(c|n)$ is the class-membership probability of n in c as determined by the probabilistic lexicon, and $f(n)$ is the frequency of n in the combined sample of objects and translation alternatives[1].

Consider example ID 160867 from Fig. 3. The ambiguity to be resolved concerns the direct objects of the verb *cross* whose lexical entry is partly shown in Fig. 2. Class 19 and the noun *border* is the pair yielding a higher estimated frequency than any other combination of a class and an alternative translation such as *boundary*. Similarly, for example ID 301946, the pair of the

---

[1] Note that $p(\hat{c}) = \max\limits_{c \in C} p(c)$ in most, but not all cases.

target-noun *society* and class 6 gives highest estimated frequency of the objects of *mobilize*.

4   Evaluation

We evaluated our resolution methods on a pseudo-disambiguation task similar to that used in Rooth et al. (1999) for evaluating clustering models. We used a test set of 298 (v, n, n') triples where (v, n) is chosen randomly from a test corpus of pairs, and n' is chosen randomly according to the marginal noun distribution for the test corpus. Precision was calculated as the number of times the disambiguation method decided for the non-random target noun ($\hat{n} = n$).

As shown in Fig. 4, we obtained 88 % precision for the class-based lexicon (ProbLex), which is a gain of 9 % over the best clustering model and a gain of 15 % over the human baseline[2].

ambiguity	human baseline	clustering	ProbLex
2	73.5 %	79.0 %	88.3 %

Figure 4: Evaluation on pseudo-disambiguation task for noun-ambiguity

The results of the pseudo-disambiguation could be confirmed in a further evaluation on a large number of randomly selected examples of a real-world bilingual corpus. The corpus consists of sentence-aligned debates of the European parliament (mlcc = multilingual corpus for cooperation) with ca. 9 million tokens for German and English. From this corpus we prepared a gold standard as follows. We gathered word-to-word translations from online-available dictionaries and eliminated German nouns for which we could not find at least two English translations in the mlcc-corpus. The resulting 35 word dictionary is shown in Fig. 5. Based on this dictionary, we extracted all bilingual sentence pairs from the corpus which included both the source-noun and the target-noun. We restricted the resulting ca. 10,000 sentence pairs to those which included a source-noun from this

---

[2]Similar results for pseudo-disambiguation were obtained for a simpler approach which avoids another EM application for probabilistic class labeling. Here $\hat{n}$ (and $\hat{c}$) was chosen such that $f_{\hat{c}}(v, \hat{n}) = \max_{c,n}((f_{LC}(v, n) + 1) p_{LC}(c|v, n))$. However, the sensitivity to class-parameters was lost in this approach.

dictionary in the object position of a verb. Furthermore, the target-object was required to be included in our dictionary and had to appear in a similar verb-object position as the source-object for an acceptable English translation of the German verb. We marked the German noun n_g in the source-sentence, its English translation n_e as appearing in the corpus, and the English lexical verb v_e. For the 35 word dictionary of Fig. 5 this semi-automatic procedure resulted in a test corpus of 1,340 examples. The average ambiguity in this test corpus is 8.63 translations per source-word. Furthermore, we took the semantically most distant translations for 25 words which occured with a certain frequency in the evaluation corpus. This gave a corpus of 814 examples with an average ambiguity of 2.83 translations. The entries belonging to this dictionary are highlighted in **bold face** in Fig. 5. The dictionaries and the related test corpora are available on the web[3].

We believe that an evaluation on these test corpora is a realistic simulation of the hard task of target-language disambiguation in real-word machine translation. The translation alternatives are selected from online dictionaries, correct translations are determined as the actual translations found in the bilingual corpus, no examples are omitted, the average ambiguity is high, and the translations are often very close to each other. In contrast to this, most other evaluations are based on frequent uses of only two clearly distant senses that were determined as interesting by the experimenters.

Fig. 6 shows the results of lexical ambiguity resolution with probabilistic lexica in comparison to simpler methods. The rows show the results for evaluations on the two corpora with average ambiguity of 8.63 and 2.83 respectively. Column 2 shows the percentage of correct translations found by disambiguation by random choice. Column 3 presents as another baseline disambiguation with the major sense, i.e., always choose the most frequent target-noun as translation of the source-noun. In column 4, the empirical distribution of (v, n) pairs in the training corpus extracted from the BNC is used as disambiguator. Note that this method yields good results in terms of precision (P = #correct / #correct + #incorrect), but is much

---

[3]http://www.ims.uni-stuttgart.de/projekte/gramotron/

Angriff	aggression, assault, offence, onset, onslaught, attack , charge, raid, whammy, inroad
Art	**form, type, way**, fashion, fit, kind, wise, manner, species, mode, sort, variety
Aufgabe	**abandonment, office, task**, exercise, lesson, giveup, job , problem, tax
Auswahl	eligibility, selection, choice, varity, assortment, extract, range, sample
Begriff	concept, item, notion, idea
Boden	**ground, land, soil**, floor, bottom
Einrichtung	**arrangement, institution**, constitution, establishment, feature, installation, construction, setup, adjustment, composition, organization
Erweiterung	**amplification, extension**, enhancement, expansion, dilatation, upgrading, add-on, increment
Fehler	**error, shortcoming**, blemish, blunder, bug, defect, demerit, failure, fault, flaw, mistake, trouble, slip, blooper, lapsus
Genehmigung	permission, approval, consent, acceptance, approbation, authorization
Geschichte	**history, story**, tale, saga, strip
Gesellschaft	**company, society**, companionship, party, associate
Grenze	**border, frontier**, boundary, **limit**, periphery, edge
Grund	**master, matter, reason**, base, cause, ground, bottom root
Karte	**card, map**, ticket, chart
Lage	site, situation, position, bearing, layer, tier
Mangel	deficiency, lack, privation, want, shortage, shortcoming, absence, dearth, demerit, desideratum, insufficiency, paucity, scarceness
Menge	**amount, deal, lot**, mass, multitude, plenty, quantity, quiverful, volume, abundance, aplenty, assemblage , crowd, batch, crop, heap, lashings, scores, set, loads, bulk
Prüfung	**examination, scrutiny, verification**, ordeal, test, trial, inspection, tryout, assay, canvass, check, inquiry, perusal, reconsideration, scruting
Schwierigkeit	**difficulty, trouble**, problem, severity, ardousness, heaviness
Seite	**page, party, side**, point, aspect
Sicherheit	**certainty, guarantee, safety**, immunity, security , collateral , doubtlessness, sureness, deposit
Stimme	**voice, vote**, tones
Termin	**date, deadline, meeting**, appointment, time, term
Verbindung	**association, contact, link**, chain, conjunction, connection, fusion, joint , compound, alliance, catenation, tie, union, bond, interface, liaison, touch, relation, incorporation
Verbot	ban, interdiction, prohibition, forbiddance
Verpflichtung	**commitment, obligation, undertaking**, duty, indebtedness , onus, debt, engagement, liability, bond
Vertrauen	**confidence, reliance, trust**, faith, assurance, dependence, private, secret
Wahl	**election, option**, choice , ballot, alternative, poll , list
Weg	**path, road, way**, alley, route, lane
Widerstand	resistance, opposition, drag
Zeichen	character, icon, sign, signal, symbol, mark, token, figure, omen
Ziel	**aim, destination, end**, designation, target, goal, object, objective, sightings, intention, prompt
Zusammenhang	coherence, context, contiguity, connection
Zustimmung	**agreement, approval, assent**, accordance, approbation, consent, affirmation, allowance, compliance, compliancy, acclamation

Figure 5: Dictionaries extracted from online resources

ambiguity	random	major sense	empirical distrib.	clustering	ProbLex
8.63	14.2 %	31.9 %	P: 46.1 % E: 36.2 %	43.3 %	49.4 %
2.83	35.9 %	45.5 %	P: 60.8 % E: 49.4 %	61.5 %	68.2 %

Figure 6: Disambiguation results for clustering versus probabilistic lexicon methods

worse in terms of effectiveness (E = #correct / #correct + #incorrect + #don't know). The reason for this is that even if the distribution of (v, n) pairs is estimated quite precisely for the pairs in the large training corpus, there are still many pairs which receive the same or no positive probability at all. These effects can be overcome by a clustering approach to disambiguation (column 5). Here the class-smoothed probability of a (v, n) pair is used to decide between alternative target-nouns. Since the clustering model assigns a more fine-grained probability to nearly every pair in its domain, there are no don't know cases for comparable precision values. However, the semantically smoothed probability of the clustering models is still too coarse-grained when compared to a disambiguation with a probabilistic lexicon. Here a further gain in precision and equally effectiveness of ca. 7 % is obtained on both corpora (column 6). We conjecture that this gain can be attributed to the combination of frequency information of the nouns and the fine-tuned distribution on the selection classes of the the nominal arguments of the verbs. We believe that including the set of translation alternatives in the ProbLex distribution is important for increasing efficiency, because it gives the disambiguation model the opportunity to choose among unseen alternatives. Furthermore, it seems that the higher precision of ProbLex can not be attributed to filling in zeroes in the empirical distribution. Rather, we speculate that ProbLex intelligently filters the empirical distribution by reducing maximal

counts for observations which do not fit into classes. This might help in cases where the empirical distribution has equal values for two alternatives.

source	target	correct	accept.
Seite	page side	76.9 %	76.9 %
Sicherheit	guarantee safety	93.8 %	93.0 %
Verbindung	connection link	58.8 %	93.8 %
Verpflichtung	commitment obligation	73.2 %	94.1 %
Ziel	objective target	72.5 %	85.5 %
overall precision		78 %	90 %

Figure 7: Precision for finding correct and acceptable translations by lexicon look-up

Fig. 7 shows the results for disambiguation with probabilistic lexica for five sample words with two translations each. For this dictionary, a test corpus of 219 sentences was extracted, 200 of which were additionally labeled with acceptable translations. Precision is 78 % for finding correct translations and 90 % for finding acceptable translations.

Furthermore, in a subset of 100 test items with average ambiguity 8.6, a human judge having access only to the English verb and the set of candidates for the target-noun, i.e. the information used by the model, selected among translations. On this set, human precision was 39 %.

5 Discussion

Fig. 8 shows a comparison of our approach to state-of-the-art unsupervised algorithms for word sense disambiguation. Column 2 shows the number of test examples used to evaluate the various approaches. The range is from ca. 100 examples to ca. 37,000 examples. Our method was evaluated on test corpora of sizes 219, 814, and 1,340. Column 3 gives the average number of senses/translations for the different disambiguation methods. Here the range of the ambiguity rate is from 2 to about 9 senses[4]. Column 4

shows the random baselines cited for the respective experiments, ranging from ca. 11 % to 50 %. Precision values are given in column 5. In order to compare these results which were computed for different ambiguity factors, we standardized the measures to an evaluation for binary ambiguity. This is achieved by calculating $p^{1/\log_2 amb}$ for precision p and ambiguity factor amb. The consistency of this "binarization" can be seen by a standardization of the different random baselines which yields a value of ca. 50 % for all approaches[5]. The standardized precision of our approach is ca. 79 % on all test corpora. The most direct point of comparison is the method of Dagan and Itai (1994) which gives 91.4 % precision (92.7 % standardized) and 62.1 % effectiveness (66.8 % standardized) on 103 test examples for target word selection in the transfer of Hebrew to English. However, compensating this high precision measure for the low effectiveness gives values comparable to our results. Dagan and Itai's (1994) method is based on a large variety of grammatical relations for verbal, nominal, and adjectival predicates, but no class-based information or slot-labeling is used. Resnik (1997) presented a disambiguation method which yields 44.3 % precision (63.8 % standardized) for a test set of 88 verb-object tokens. His approach is comparable to ours in terms of informedness of the disambiguator. He also uses a class-based selection measure, but based on WordNet classes. However, the task of his evaluation was to select WordNet-senses for the objects rather than the objects themselves, so the results cannot be compared directly. The same is true for the SENSEVAL evaluation exercise (Kilgarriff and Rosenzweig, 2000)—there word senses from the HECTOR-dictionary had to be disambiguated. The precision results for the ten unsupervised systems taking part in the competitive evaluation ranged from 20-65% at efficiency values from 3-54%. The SENSEVAL standard is clearly beaten by the earlier results of Yarowsky (1995) (96.5 % precision) and Schütze (1992) (92 % precision). However, a comparison to these re-

---

[4]The ambiguity factor 2.27 attributed to Dagan and Itai's (1994) experiment is calculated by dividing their average of 3.27 alternative translations by their average of 1.44 correct translations. Furthermore, we calculated the ambiguity factor 3.51 for Resnik's (1997) experiment

from his random baseline 28.5 % by taking 100/28.5; reversely, Dagan and Itai's (1994) random baseline can be calculated as 100/2.27 = 44.05. The ambiguity factor for SENSEVAL is calculated for the noun task in the English SENSEVAL test set.

[5]Note that we are guaranteed to get exactly 50 % standardized random baseline if $random \cdot amb = 100$ %.

disambiguation method	corpus size	ambiguity	random	precision	random (standardized)	precision (standardized)
ProbLex	1 340	8.63	14.2 %	49.4 %	53.4 %	79.7 %
	814	2.83	35.9 %	68.2 %	50.5 %	77.5 %
	219	2	50.0 %	78.0 %	50.0 %	78.0 %
Dagan, Itai 94	103	2.27	44.1 %	P: 91.4 % E: 62.1 %	50.0 %	P: 92.7 % E: 66.8 %
Resnik 97	88	3.51	28.5 %	44.3 %	50.0 %	63.8 %
SENSEVAL 00	2 756	9.17	10.9 %	P: 20-65 % E: 3-54 %	50.0 %	P: 60-87 % E: 33-83 %
Yarowsky 95	∼ 37 000	2	50.0 %	96.5 %	50.0 %	96.5 %
Schütze 92	∼ 3 000	2	50.0 %	92.0 %	50.0 %	92.0 %

Figure 8: Comparison of unsupervised lexical disambiguation methods.

sults is again somewhat difficult. Firstly, these approaches were evaluated on words with two clearly distant senses which were determined by the experimenters. In contrast, our method was evalutated on randomly selected actual translations of a large bilingual corpus. Furthermore, these approaches use large amounts of information in terms of linguistic categorizations, large context windows, or even manual intervention such as initial sense seeding (Yarowsky, 1995). Such information is easily obtainable, e.g., in IR applications, but often burdensome to gather or simply unavailable in situations such as incremental parsing or translation.

6 Conclusion

The disambiguation method presented in this paper deliberately is restricted to the limited amount of information provided by a probabilistic class-based lexicon. This information yet proves itself accurate enough to yield good empirical results, e.g., in target-language disambiguation. The probabilistic class-based lexica are induced in an unsupervised manner from large unannotated corpora. Once the lexica are constructed, lexical ambiguity resolution can be done by a simple lexicon look-up. In target-word selection, the most frequent target-noun whose semantics fits best to the semantics of the argument-slot of the target-verb is chosen. We evaluated our method on randomly selected examples from real-world bilingual corpora which constitutes a realistic hard task. Disambiguation based on probabilistic lexica performed satisfactory for this task. The lesson learned from our experimental results is that hybrid models combining frequency information and class-based

probabilities outperform both pure frequency-based models and pure clustering models. Further improvements are to be expected from extended lexica including, e.g., adjectival and prepositional predicates.

References

Glenn Carroll and Mats Rooth. 1998. Valence induction with a head-lexicalized PCFG. In *Proceedings of EMNLP-3*, Granada.

Ido Dagan and Alon Itai. 1994. Word sense disambiguation using a second language monolingual corpus. *Computational Linguistics*, 20:563–596.

Adam Kilgarriff and Joseph Rosenzweig. 2000. English SENSEVAL: Report and results. In *Proceedings of LREC 2000*.

Philip Resnik. 1997. Selectional preference and sense disambiguation. In *Proceedings of the ANLP'97 Workshop: Tagging Text with Lexical Semantics: Why, What, and How?*, Washington, D.C.

Mats Rooth, Stefan Riezler, Detlef Prescher, Glenn Carroll, and Franz Beil. 1999. Inducing a semantically annotated lexicon via EM-based clustering. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL'99)*, Maryland.

Hinrich Schütze. 1992. Dimensions of meaning. In *Proceedings of Supercomputing '92*.

David Yarowsky. 1995. Unsupervised word sense disambiguation rivaling supervised methods. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics (ACL'95)*, Cambridge, MA.