CENTER FOR INFORMATION AND LANGUAGE PROCESSING

Supervised Word Disambiguation for Venetan: a proof-of-concept experiments

Final Report of the participation to FLAIRS 2017
30*th* Florida Artificial Intelligence Research Society International Conference

Student: Costanza Conforti. Supervisor: Dr Alexander Fraser
May 30, 2017

1 OVERVIEW OF THE RESEARCH PROJECT

- **Motivation**

  In this project, we investigated a well-known Natural Language Processing (NLP) task, Word Sense Disambiguation (WSD). WSD is a classification task that consists of determining which of the senses of an ambiguous word is activated in a specific context. Research in this field has primarily concentrated on investigating English and a few other well-resourced languages. Goal of this project was to examine to which extend traditional Machine Learning methods, as Support Vector Machines, could perform on a very small corpus. In particular, we worked with Venetan, an Italian dialect spoken in the northern part of Italy. Venetan is a widely studied language from a linguistic point of view. However, it lacks of resources and tools for NLP. Moreover, inconsistencies in the orthography and strong morphological variations constitute a challenge for developing NLP system dealing with Venetan.

  Our main contributions are twofold: first, we select and annotate a corpus for Venetan, considering two words (one abstract and one concrete term) and using two levels of annotation (fine- and coarse-grained), reporting on annotator agreement. Second, we report results of proof-of-concept experiments of supervised WSD performed with Support Vector Machines on this corpus. To our knowledge, our work is the first time that WSD for a European Dialect like Venetan has been studied.

- **Results**

  For a detailed presentation and discussion of methods and results, please see the paper and the poster included. Two native speakers annotated the data. The labelling task has been defined as a discrimination task. Overall, we implemented six features and we carried on experiments concerning two words and two level of annotation (fine- and coarse-grained). Results have been compared with upper and lower bounds. As baselines, two dummy classifiers were considered, while judges agreement rate was taken as the upper bound. As the amount of data available was scarce, a 10-fold cross validation strategy. We observed that different feature combinations performed better for a specific word, with the best feature combinations reaching a performance few points below the upper bound.

  To conclude, our proof-of-concept results are promising and demonstrate that, even with limited resources, the problem of WSD for an Italian dialect can be concretely approached.

2 ATTENDING FLAIRS

In November 2016, we submitted our project to the 30*th* Florida Artificial Intelligence Research Society (FLAIRS) International Conference. FLAIRS is a renowned international conference on Artificial Intelligence, which reunites well-known speakers from fields like Data Mining, Big Data analysis, Recommender Systems, Natural Language Processing and many others.

In March 2017, we have been notified that our work had been accepted as a short paper at FLAIRS, and that it had been presented as a poster. The paper will appear in the Proceedings of the Conference, which are published by the Association for the Advancement of Artificial Intelligence Press (AAAI).

30*th* FLAIRS International Conference took place on May 22-24, 2017 in Florida. Costanza Conforti participated to the whole conference and presented our project on May 22. During the poster presentation, she had the possibility to get in contact with many researchers working in the field of Natural Language Processing in both Europe and US, having the opportunity to collect valuable suggestions about our paper and to establish useful connections with researchers from other Universities. Moreover, attending to other talks and presentations gave her valuable overviews on interesting Machine Learning applications both inside and outside the NLP field. The paper and the poster are included as appendices to this report.

3 CONCLUSIONS

This research has been generously supported by Lehre@LMU. We are extremely grateful to the Lehre@LMU committee for selecting our project and for financing my participation to the conference. Having the opportunity to present our work in such a prestigious location was extremely fruitful from an academic point of view, and we are proud of having represented LMU at FLAIRS 2017.

Supervised Word Sense Disambiguation for Venetan: a Proof-of-Concept Experiment

Costanza Conforti and **Alexander Fraser**
Center for Information and Language Processing
Ludwig-Maximilians University of Munich
Oettingenstrae 67, 80583 - Munich, Germany

Abstract

Word Sense Disambiguation (WSD) is a classification task that consists of determining which of the senses of an ambiguous word is activated in a specific context. Research in this field has primarily concentrated on investigating English and a few other well-resourced languages. Recently, studies done on a corpus of Old English (Wunderlich 2015) showed that, even with limited resources, it is still possible to approach the problem of WSD. In this paper a WSD system has been developed for the Low Resource Language (LRL) Venetan, which has recently received some attention from the Natural Language Processing (NLP) community. Our main contributions are twofold: first, we select and annotate a corpus for Venetan, considering two words (one abstract and one concrete term) and using two levels of annotation (fine- and coarse-grained), reporting on annotator agreement. Second, we report results of proof-of-concept experiments of supervised WSD performed with Support Vector Machines on this corpus. To our knowledge, our work is the first time that WSD for a European Dialect like Venetan has been studied.

Introduction

NLP research in the field of LRLs is hindered by a number of factors, like the lack of lexicographical resources, combined with the scarcity of available and labeled data. NLP for Italian dialects[1] has not been investigated much. Even if they are usually relatively close to the standard language, adapting existing tool for Italian can be very challenging, as dialects are very fragmented and show many orthographic and morphological variants. This constitutes a general challenge in NLP applications focusing on LRL and Old Languages. Aside from a few contributions (Bortolotti 2005), work on Italian dialects mainly concentrated on Venetan, a LRL primarily spoken in the Northeastern regions of Italy. The term LRL can be used to refer to a number of different situations. Following the classification proposed by Singh (2008), Venetan can be considered a resource scarce language, as it has been widely studied from a linguistic point of view, but very few resources and tools for NLP are available. In recent years, the research community started to investigate some classic NLP tasks like morphological analysis (Tonelli 2010) and POS tagging, (Jaber 2011), including a preliminary study on Statistical Machine Translation from Venetan into English (Delmonte 2009).

A brief description of Venetan

Venetan is a Gallo-Italic language spoken in the Northeast region of Italy, where it is spoken as mother tongue by about 3.852.500 people. Ethnologue reports also 50000 speakers in Croatia and about 4 million in Brazil, where this language is called *Talian*[2]. Venetan is in vigorous use: proficiency among local speakers reaches 75% of the population and it is widely spoken within generations (Tonelli 2010). A number of newspapers are published partially in Venetan and some radio stations broadcast in Venetan; moreover, a small community of Venetans is very active on the Internet, where many blogs and websites can be found, including a version of Wikipedia[3]. Far from constituting a standardized language, Venetan constitutes a regional continuum. Linguists have identified at least four main varieties of Venetan, which show peculiarities in morphology, syntax and lexicon but are still mutually comprehensible. Therefore, Venetan can be considered as a diasystem, where speakers use their own variety in everyday life and manage to understand each other (Tonelli 2010). Venetan is usually written with the Italian alphabet, plus some special characters. Some attempts to unify orthography have been made, like the project promoted by the Veneto Region that culminated in the *Manual of Venetian Orthography*. However, these guidelines are far from being universally accepted and inconsistencies in the orthography and strong morphological variations are both frequent.

Corpus creation

In this Section, we describe the linguistic resources and the preprocessing pipeline used to extract candidates for WSD and for annotating the training samples.

[1] In this paper, we use the word *language* and *dialect* indiscriminately. Partly autonomous, Italian dialects came into being more or less in the same period through transformations of Latin. Many centuries later, one of them, the dialect of Florence, became the official language of the Italian State in 1861. However, from a historical and linguistic point of view, each dialect can be seen as a language on its own (Berruto 2005).

[2] https://www.ethnologue.com/language/vec

[3] https://vec.wikipedia.org/wiki/V%C3%A8neto

Resources Description

We used the STILVEN corpus as our source for extracting training examples. STILVEN is a project founded by the Veneto Region in 2008[4] and carried out by researchers from the Ca' Foscari University (Venice) and IRST-FBK (Trento). The project involved the implementation of a morphological analyzer for Venetan which permitted the creation of a corpus with a homogenized orthography (Tonelli 2010), and the development of a Venetan POS-tagger (Jaber 2011). To our knowledge, STILVEN is the only available corpus for an Italian dialect which is orthographically normalized and POS-tagged. The corpus collects very heterogeneous texts, including children stories, famous quotes, a manual of Venetan orthography rules and translations of a book about American history and of *The Little Prince* by Antoine de Saint-Exupry. Statistical information about the STILVEN corpus are listed in Table 1.

Token count	133734
Type count	13321
Ratio of (token count/type count)	10.03
Total number of sentences	13058
Average sentence length	10.24
Minimum sentence length	2
Maximum sentence length	77

Table 1: STILVEN Corpus statistics

Candidate Extraction

In order to obtain a list of polysemous words, we searched in the STILVEN corpus for common nouns with minimum token frequency of 95 and minimum word length of 3 characters[5]. Seven words meet these criteria. All seven terms were, at different levels, polysemous[6] (Table 2). In the following sections, the annotation process and WSD results for the words parte (Engl. *part*, *somewhere*) and *omeni* (Engl. *men*, *soldiers*) are analysed in detail. Concerning the occurrences of these two words, we excluded the sentences where the candidate was part of a collocation (as for example *omeni de afari*, Engl. *businessmen*). Three occurrences of *parte* were also discarded, as they had been wrongly tagged as noun while they were actually a verb (Eng. *leave*). The final occurrence count was of 98 occurrences for the word *omeni* and 121 for *parte*.

[4]project.cgm.unive.it/stilven_en.html

[5]We considered only words occurring with a frequency higher than 95 in order to obtain a corpus sufficiently large to be used for training a WSD system. This methodology is similar to the one applied in (Wunderlich 2015).

[6]However, three words in this list have been discarded due to different reasons. The sense distributions of idea (Engl. *idea*) and man (Engl. *help, hand*) were too unbalanced (90/11/106 considering the three senses of *idea*, 113/23 considering the two senses of *man*). The word *dito*, (Engl. *finger, proverb*) had been tagged as a noun, but in more then a half of the occurrences it was actually the past participle of the verb *say*.

Word Token	Count	English Translation
dito	251	*finger, proverb*
idea	208	*idea*
roba	208	*stuff, thing, food*
man	138	*help, hand*
parte	144	*part, somewhere*
tenpo	113	*time*
omeni	99	*men, soldiers*

Table 2: Candidates for WSD from the STILVEN corpus

Synset Definition and Candidate Annotation

To generate training data for supervised classification, all occurrences of *parte* and *omeni* were manually labeled with the sense activated in the sentence. To our knowledge, no lexicographic resource is available for Venetan. Therefore, we proceeded as follows: first of all, we looked up in the Venetan-Italian translation dictionary *El Galepin*[7] for the Italian translations of each Venetan word. Then, the Italian and English WordNet[8] were consulted in order to collect the *synsets* of each translation. With the help of a Venetan native speaker, we merged the most related and fine-grained *synsets* in order to obtain a small final set of clearly distinct senses. We selected three senses for the word *omeni*, while for the word *parte*, we consider three coarse-grained senses and six fine-grained senses (see Table 3). Finally, the set of senses of each word, with the related corpus, has been given to two non-professional annotators separately[9].

The effort of manual annotation was considerable, as sense annotation is a very difficult undertaking (Wilks 1998). As the annotation sessions with the second judge were quite laborious, we redefined the task from a classification to a discrimination task, as was done by Gale (1992): the second annotator received a sentence labeled by the first judge and had to report whether she agreed or not with the classification. For the final annotation, when the annotators disagreed, the label proposed by the first judge was taken as the gold standard. As reported in Table 3, the sense distributions of *omeni* and *parte* with coarse-grained labels are quite unbalanced, whereas the fine-grained classification of *parte* is more uniform. Judges' agreement for *parte* using fine-grained labels was lower than using coarse-grained labels ($\sim$79.0% vs. $\sim$93.9%). Judges' agreement for the concrete word *omeni* was lower then for the abstract term *parte* ($\sim$89.7%). These values are far from the 96.8% agreement obtained by Gale (1992), but this difference can be explained considering that the annotators were not professionals.

Methods and Evaluation Metrics

Supervised WSD was performed on the annotated corpora using Support Vector Machines (SVMs). SVMs were cho-

[7]http://www.elgalepin.com/. The dictionary has been released online in 2007 and counts around 37.000 entries.

[8]We used the MultiWordNets on-line interface available at http://multiwordnet.fbk.eu/english/home.php (Artale 1997)

[9]The annotators were a 23 and a 54-year-old women, with no special linguistic training. Both are native speakers of Venetan.

Sense	%
Omeni	
1. Adult male person (opposed to woman)	**14**
2. A human being	**59**
3. Soldier	**27**
***Parte* with coarse-grained labels**	
1. Something less than the whole	**61**
2. Role	**8**
3. Road or path (generic)	**31**
***Parte* with fine-grained labels**	
1a. Something determined in relation to an entity	**31**
1b. Region or state	**9**
1c. One of the portions into which something is divided and which together constitute a whole	**21**
2. Role	**8**
3a. A line leading to a place or point	**18**
3b. Space for movement	**13**

Table 3: Sense labels and distribution

sen for a number of reasons: first of all, according to Navigli (2009), they achieve the best results in WSD compared to several other supervised methods. In particular, SVMs work efficiently in environments where there are a large number of features (Cabezas 2001) and are usually more resistant to overfitting (Lee 2004). Moreover, as stated in Yarowsky (2010), SVMs often perform well with few training examples per label. We used the implementation of LinearSVM in scikit-learn[10]. In the following subsections, the extracted features and the evaluation metrics are described.

Feature Design

For designing features, we mainly followed Lee (2004) and Cabezas (2001). Overall, six features were implemented:

- Unordered bag-of-words (BoW) vector, considering all the lowercased words in the sentence.
- Unordered BoW vector with stopwords removed (as done by Lee 2004). We obtained a *stopword* list by selecting the most common tokens whose POS tag was in a restricted list (including articles, pronouns, clitics and conjunctions).
- Unordered BoW vector considering the wide-context (two sentences preceding and following the occurrence), which has been proved to improve noun disambiguation (Yarowsky 2010).
- Unordered BoW vector of the wide-context after filtering out *stopwords*.
- POS tags of the three words preceding and following the occurrence. The POS tag of the null token was denoted with a special symbol.
- Ordered sequence of tokens in the local, narrow context of the occurrence. Following Lee (2004), 11 features were developed, corresponding to different collocations.

[10] http://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVC.html

Evaluation Metrics

For each feature combination, the following evaluation metrics were calculated using scikit-learn: Accuracy, Precision, Recall and balanced F1 Measure. These metrics were compared with upper and lower bounds. As baselines, two dummy classifiers were considered: the first one performs random classification, whereas the second one always chooses the most common class. Judges agreement rate was taken as the upper bound.

Experiments and Results

We performed experiments using different feature combinations. As the amount of data available was scarce, a 10-fold cross validation strategy was used. In general, WSD tasks are difficult due to the very unbalanced sense distribution. Results with different feature combinations are reported in Table 4. Considering the word *omeni*, the most informative feature combination was BoW with the wide-context and the POS feature. As shown in Table 5, using this combination returns lower Recall for the first sense of *omeni*, but considerably improves all measures for the third sense, which was the lowest represented. In fact, occurrences of this sense usually occur in texts about American History, for which the wide-context feature can be useful in disambiguation. On the contrary, first sense occurrences often appear in quotations, which are unrelated to each other and for which the wide-context feature can be misleading.

Moving to the word *parte*, classification with coarse- or fine-grained labels follows the same pattern. In contrast to what happened with *omeni*, considering the wide context has a negative effect on the overall accuracy, whereas information about POS tags seems to be useful (see Table 4). However, best results are achieved using only local collocations together with the BoW feature. In fact, the corpus for the word *parte* is highly repetitive, so that considering ordered sequences of words near to the occurrence, like in the local collocations feature, can be very useful for disambiguation. As shown in Table 5, using this feature combination leads to acceptable Recall even for the second sense, which had very few occurrences.

	omeni	***parte I***	***parte II***
Random classifier	0.43	0.32	0.10
Most common classifier	0.60	0.64	0.32
BoW	0.77	0.81	0.58
Bow + wide_context	0.81	0.80	0.57
BoW+POS	0.73	0.85	0.62
BoW+wide_context+POS	**0.82**	0.85	0.62
BoW+collocations	0.75	**0.89**	**0.70**
BoW+collocations+POS	0.81	0.88	0.70
Judges agreement	0.89	0.93	0.79

Table 4: WSD Accuracy results using different feature combinations for *omeni*, *parte* with coarse labels (I) and *parte* with fine grained labels (II), compared with the baselines and upper bound.

	1st sense		2nd sense		3rd sense	
	P	R	P	R	P	R
omeni						
baseline	0.62	0.36	0.76	0.95	0.89	0.62
best comb	0.67	0.29	0.80	0.97	0.95	0.81
parte I						
baseline	0.86	0.92	0.00	0.00	0.76	0.82
best comb	0.89	0.97	0.71	0.56	0.94	0.82

Table 5: Comparison of Precision and Recall using only BoW feature and with the best combination. Results for *parte* with fine-grained labels follow the same pattern and are not reported due to lack of space.

Conclusions and Future Work

In this paper, we reported on our annotation of a gold-standard and on the results of supervised WSD considering two Venetan words.

The annotation phase was laborious and time-consuming. As we were dealing with a LRL, we were not able to find expert annotators of Venetan, such as lexicographers (as in Wilks (1998)). Following Gale (1992), the difficulties of working with non-professional annotators were partially solved by redefining word-sense classification to a discrimination task. Concerning the results of WSD, we observed that different feature combinations performed better for a specific word. This is consistent with Resnik's statement (Resnik 1997), according to which disambiguation, as a highly lexically sensitive task, in effect requires a specialized disambiguator for each considered word. In contrast with works on WSD for Old Languages and for other LRLs, (e.g., (Wunderlich 2015)), in our work we were able to access information from POS-tags. But, contrary to what we expected, considering POS-tags was not decisive for improving disambiguation. Furthermore, filtering out stopwords from the BoW features was not helpful for disambiguation. This experiment could be repeated using a different strategy to obtain a *stopwords* list. Additional knowledge sources could further improve accuracy if more tools for Venetan become available in the future. Future work could also investigate the adaptation of existing NLP tools for Italian to Venetan, as the two languages show a high degree of similarity (Tonelli 2010). It could be particularly interesting to obtain information about syntactic relations, which have been shown to be very discriminative in WSD (Lee 2004).

Overall, our proof-of-concept results are promising and demonstrate that, even with limited resources, the problem of WSD for an Italian dialect can be concretely approached, and we hope that our work will encourage further work on European dialects.

Aknowledgments

The authors are thankful to Prof Rodolfo Delmonte of the Ca' Foscari University of Venice for providing us with the STILVEN corpus.

References

Artale, A., Magnini, B., and Strapparava, C. 1997. Lexical discrimination with the Italian version of WordNet. In *ACL Workshop on Information Extraction*

Berruto, G. 2005. Dialect/standard convergence, mixing, and models of language contact: the case of Italy. *Dialect change. Convergence and divergence in European languages*, 81-97.

Bortolotti E., Rasom, S. 2005. The Ladin language between fragmentation and standardization: the contribution of Computational Linguistics, *Lesser Used Languages and Computer Linguistics, LULCL*

Cabezas, C., Resnik, P., and Stevens, J. 2001. Supervised sense tagging using support vector machines. In *The Proceedings of the Second International Workshop on Evaluating Word Sense Disambiguation Systems*: 59-62. Association for Computational Linguistics

Delmonte, R., Bristot, A., Tonelli, S., and Pianta, E. 2009. English/Veneto resource poor machine translation with STILVEN. In *Proceedings of the International Symposium on Data and Sense Mining, Machine Translation and Controlled Languages (ISMTCL)*

Gale, W., Church, K., and Yarowsky, D. 1992. Estimating upper and lower bounds on the performance of word-sense disambiguation programs. In *Proceedings of the 30th annual meeting on Association for Computational Linguistics*: 249-256. Association for Computational Linguistics

Jaber, S., Tonelli, S., and Delmonte, R. (2011). Venetan to English Machine Translation: Issues and Possible Solutions. In *8th International NLPCS Workshop*: 69 ISO 690

Lee, Y. K., Ng, H. T., and Chia, T. K. 2004. Supervised word sense disambiguation with support vector machines and multiple knowledge sources. In *Senseval-3: third international workshop on the evaluation of systems for the semantic analysis of text*: 137-140

Navigli, R. 2009. Word sense disambiguation: A survey. *ACM Computing Surveys CSUR* 41(2), 10. ISO 690

Resnik, P., and Yarowsky, D. 1997. A perspective on word sense disambiguation methods and their evaluation. In *Proceedings of the ACL SIGLEX workshop on tagging text with lexical semantics: Why, what, and how*: 79-86

Singh, A. K. 2008. Natural Language Processing for Less Privileged Languages: Where do we come from? Where are we going? In *IJCNLP*: 7-12

Tonelli, S., Pianta, E., Delmonte, R., and Brunelli, M. 2010. VenPro: A Morphological Analyzer for Venetan. In *LREC*

Wilks, Y., and Stevenson, M. 1998. Word sense disambiguation using optimised combinations of knowledge sources. In *Proceedings of the 17th international conference on Computational linguistics* (2):1398-1402. Association for Computational Linguistics.

Wunderlich, M., Fraser, A., and Langeslag, P. S. 2015. God Wat Thaet Ic Eom God - An Exploratory Investigation Into Word Sense Disambiguation in Old English. In *GSCL*: 39-48

Yarowsky, D. 2010. Word sense disambiguation. In *Handbook of Natural Language Processing*, Second Edition 315-338. Chapman and Hall/CRC.

Supervised Word Sense Disambiguation for Venetan: a Proof-of-Concept Experiment

Costanza Conforti and Alexander Fraser

Center for Information and Language Processing

Ludwig-Maximilians University of Munich

This research has been partially funded by Lehre@LMU

Overview

- **Word Sense Disambiguation** (WSD) : classification task that consists of determining which of the senses of an ambiguous word is activated in a specific context (Navigli 2009)

- **Venetan :**
  - ~8M speakers in Italy, Brazil and Croatia
  - widely studied from a linguistic point of view
  - challenges for NLP: inconsistencies in orthography, strong morphological variations within different varieties

- **Resources :**
1. **STILVEN corpus** (Delmonte 2009)
   - Ca' Foscari University, IRST-FBK
   - ~140k tokens from very heterogeneous sources
   - homogenized orthography, POS tagged
2. ***El Galepin* translation dictionary**

Data

1. Candidate Extraction
- common polysemic nouns
- token frequency > 95

→

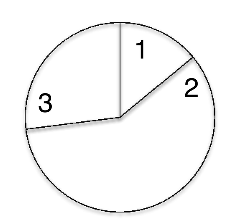
2. Synset Definition
- Venetan words translated into Italian using *Galepín*
- lookup into the Italian WordNet
- *synsets* refinement

→

3. Corpus Annotation
- 2 non-professional native speaker annotators
- judges agreement: 79/93%
- annotation as discrimination task (Gale 1992)

Results reported on 2 words:

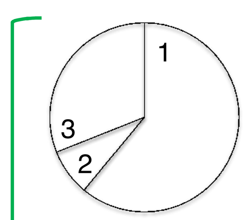
1. òmeni
- concrete word
- 3 senses: *men, mankind, soldiers*

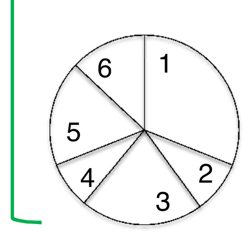


1. `xe raro ke a łe done ghe piaxa i òmeni ke i ga masa rispeto pa' łore .`
   it is rare that men who nurture the utmost respect for women are held in high regard by them

2. parte
- abstract word
- 2 levels of classification:

- **coarse-grained *(parte I):*** 3 senses: *part, role, somewhere*
- **fine-grained *(parte II):*** 6 senses: *part (as part#1*), region (as part#5*), portion (as part#2*), role, path, somewhere*





2. `i òmeni i fa mal , Come ke łe ave łe fa el miel .`
   man produces evil as a bee produces honey

3. `inte el 1758 el Massachùssetts el ga rancurà su 7200 òmeni , Connecticut ghin' à armà 'sincoe-miła' […]`
   In 1758 Massachusetts raised an army of 7,200 soldiers , Connecticut put 5,000 under arms […]

* Refers to the English Wordnet Sense Number

Methods

- **Support Vector Machines (SVM)**
  - best results in WSD compared to several other supervised methods (Navigli 2009)
  - efficient in environments with large number of features, usually resistant to overfitting
  - perform well even with few training examples per class (Lee 2004)

- **Features** (mainly following Cabezas 2001 and Lee 2004)
1. Unordered bag-of-words (BoW) vector
2. Unordered BoW vector without *stopwords* (stopword list obtained by selecting the most common tokens whose POS tag was in a restricted list)
3. Unordered BoW vector considering the wide-context (2 sentences around the occurrence)
4. Unordered BoW vector of the wide-context without *stopwords*.
5. Ordered POS tags sequence of the three words around the occurrence
6. Ordered sequence of tokens in the local, narrow context of the occurrence - 11 features corresponding to different collocations

Experimental Results

Feature combination	Accuracy *òmeni*	Accuracy *parte I*	Accuracy *parte II*
I. Random classifier	0.43	0.32	0.10
II. Most common classifier	0.60	0.64	0.32
Bag-of-words (BoW)	0.77	0.81	0.58
Bow + wide context	0.81	0.80	0.57
BoW+POS	0.73	0.85	0.62
BoW+wide context+POS	**0.82**	0.85	0.62
BoW+collocations	0.75	**0.89**	**0.70**
BoW+collocations+POS	0.81	0.88	0.70
Judges agreement	0.89	0.93	0.79

baselines

maybe due to imperfect POS tagger accuracy (Delmonte 2009)

mostly in narrative sequences, where wide context can be useful

very unbalanced label distribution

Fine- and coarse-grained classification follow the same pattern.

- Different feature combinations performed better for a specific word
  → WSD, as a highly lexically sensitive task, requires a specialized disambiguator for each considered word (Resnik 1997)
- Filtering out stopwords from the BoW features was not helpful → repeat using different strategy to obtain stopword list
- Even with limited resources, the problem of WSD for a low-resources language can be approached (Wunderlich 2015)

		1st sense		2nd sense		3rd sense	
		prec	rec	prec	rec	prec	rec
òmeni	*BoW feature*	0.62	0.36	0.76	0.95	0.89	0.62
	Best comb	0.67	0.29	0.80	0.97	0.95	0.81
parte I	*BoW feature*	0.86	0.92	0.00	0.00	0.76	0.82
	Best comb	0.89	0.97	0.71	0.56	0.94	0.82

very low support for this class

Cabezas, C. *et al.* 2001. Supervised sense tagging using SVM. In *Proceedings of the 2nd International Workshop on Evaluating WSD Systems*: 59-62. ACL

Delmonte, R. *et al.* 2009. English/Veneto resource poor MT with STILVEN. In *Proceedings of the International Symposium on Data and Sense Mining*,

Gale, W. *et al.* 1992. Estimating upper and lower bounds on the performance of WSD programs. In *Proceedings of the 30th annual meeting on ACL*: 249-256

Lee, Y. K. *et al.* 2004. Supervised WSD with SVM and multiple knowledge sources. *Senseval-3*: 137-140

Navigli, R. 2009. WSD: A survey. ACM Computing Surveys CSUR 41(2)

Resnik, P., and Yarowsky, D. 1997. A perspective on WSD methods and their evaluation. *In Proceedings of the ACL SIGLEX workshop:* 79-86

Wunderlich, M. *et al.* S. 2015. God Wat Thaet Ic Eom God - An Exploratory Investigation Into WSD in Old English. In GSCL: 39-48