# Predicting Pronoun Translation Using Syntactic, Morphological and Contextual Features from Parallel Data

**Sharid Loáiciga**
Département de Linguistique
Centre Universitaire d'Informatique
Université de Genève
sharid.loaiciga@unige.ch

## Abstract

We describe the systems submitted to the shared task on pronoun prediction organized within the Second DiscoMT Workshop. The systems are trained on linguistically motivated features extracted from both sides of an English-French parallel corpus and their parses. We have used a parser that integrates morphological disambiguation and which handles the REPLACE_XX placeholders explicitly. In particular, we compare the relevance of three groups of features: a) syntactic (from the English parse), b) morphological (from the French morphological analysis) and c) contextual (from the French sentence) for French pronoun prediction. A discussion on the role of these sets of features for each pronoun class is included.

## 1 Introduction

In this paper we describe the Geneva 1 and Geneva 2 systems submitted for the shared task on pronoun prediction organized in conjunction with the EMNLP 2015 Second Workshop on Discourse in Machine Translation (MT) (Hardmeier et al., 2015). Additionally, two contrastive systems are included.

Pronouns are economical, short and independent words which can stand in the place of a more cumbersome word, and thus they lack some informativity. Their main purpose is to avoid unnecessary repetition of concepts (De Beaugrande and Dressler, 1981). Because they "cannot be interpreted without considering the discourse context", some of them are considered *anaphora* (Stede, 2012, 41). In other words, they *corefer* with other element to find their meaning.

The task of finding the referent or *the antecedent* for each anaphor is known as *Anaphora Resolution* (AR). Research on this problem has been active for some time now (Mitkov, 2001; Mitkov, 2002; Strube, 2007; Stoyanov et al., 2009; Ng, 2010). However, the independent development of MT, and Statistical Machine Translation (SMT) especially, has encountered a new dimension of the same problem: inaccurate pronoun translation. Indeed, inaccurate pronoun translation is the result of non-existent AR when passing from the source to the target language. However, plugging a AR system into the MT system has not proved to be a suitable solution to the problem (Hardmeier and Federico, 2010; Le Nagard and Koehn, 2010; Guillou, 2012). AR systems rely on a heavy preprocessing of the text, with several sub-tasks which are themselves imperfect and hard. Besides, their quality is not good enough yet to have a serious impact in MT output quality. Last, most of them exist only for English (Mitkov and Barbu, 2002; Stede, 2012).

The systems described in this paper are not developed nor intended as AR systems. Therefore, they do not explicitly search the antecedent of pronouns, but their purpose is to predict directly a pronoun translation using a classifier fed with features extracted from parallel data. They represent an alternative to the use of an AR system for helping MT. Unlike SMT systems, these classifiers have access to both source and target language data (excepting the target pronoun) during training and testing time. This data can be analyzed in order to create features which encode different types of information. Other than *generating* a possible translation, a pronoun predictor *chooses* a translation among a list of several classes.

## 2 Related Work

The idea of using word-aligned parallel data for AR was first introduced by Mitkov and Barbu (2002) to tackle difficult cases for common English AR systems. As an illustration, one of



*Proceedings of the Second Workshop on Discourse in Machine Translation (DiscoMT)*, pages 78–85,
Lisbon, Portugal, 17 September 2015. ©2015 Association for Computational Linguistics.

their examples is repeated here:

(1) a. *en* John removes the cassette from the videoplayer and disconnects **it**.
b. *fr* Jean éjecte la cassette du magnétoscope et **le** débranche.

In (1a), the pronoun *it* has both *cassette* and *videoplayer* as potential antecedents. However, the first is more prominent (a direct object), while the actual antecedent is a prepositional phrase, a syntactic type heavily penalized by most AR systems. This case can be disambiguated by looking at its gender-marked translation (1b). Since both *magnétoscope* and *le* are masculine in French, they can be matched safely as coreferring, excluding *cassette* which is feminine.

Pronoun prediction is based on the parallel data used for building SMT systems and follows Mitkov and Barbu's intuition of disambiguating pronouns based on their translation.

Building a predictor of target-language translations is a strategy introduced by Popescu-Belis et al. (2012). Using English-French parallel data, the authors manually gathered a corpus of 400 instances of *it* and their translation and used it as training data. Features include the gender of the previous ten NPs, and positional and grammatical information about the pronoun. Accuracy is reported to be around 60%.

Hardmeier, Tiedemann, and Nivre (2013) run classifiers for the same task using all the parallel data from SMT training. Their features come from the context of the pronoun in the source language (three words before and after) and from the potential antecedents (determined using a AR toolkit in the target language). In experiments with a Maximum Entropy classifier, a performance of 0.54 precision, 0.06 recall is obtained. A second set of experiments, where the AR results are dropped and a neural network classifier is used, they report precision of 0.565 and recall of 0.116. It is argued that performance is particularly good with low-frequency classes such as the feminine pronoun *elles*. In a later stage of this work, the neural network classifier is combined with a SMT system built using the Docent decoder (Hardmeier, 2014). Similarly, Weiner (2014) uses Discriminative Word Lexicon (DWL)[1] with an AR algorithm on the English side of a parallel and their correspondent word-aligned German token.

Finally, Novák (2011), Novák et al. (2013) model pronoun prediction for Czech. Features are extracted exclusively from the source text (English) following the Czech grammar rules that disambiguate the possible translations of *it*. Accuracy is around 70%.

## 3 Pronoun Mapping Between English and French

The shared task consisted in predicting the French translations of the English third-person subject pronouns *it* and *they* (Hardmeier et al., 2015). The nine classes shown in Table 1 were defined. They are presented along with their possible translations. These correspondences were determined using the word alignments provided with the training data and corrected by hand[2]. The important imbalance in the distribution of the classes, concerning the OTHER class in particular, is to be noted. A manual review of the data uncovered that this class includes translations as lexical NPs (2), other pronouns (3) and nothing at all as in the case of paraphrases (4). Object pronouns are included as well (4), (5). This is likely a source of errors, since they are homographic to subject pronouns in English.

(2) a. Certainly **it** is perceived de facto to be impossible.
b. **La chose** est certainement perçue de facto comme étant impossible.

(3) a. It was not able to do very much but **it** was repeatedly abused by Members of this House [...].
b. Elle ne permettait pas de faire grand-chose mais les députés de cette Assemblée **en** abusaient constamment [...].

(4) a. I believe **it** to be of vital importance that where Member States allow regions and local authorities to raise taxes, **they** should continue to be able to do so and not be subject to across-the-board regulation by Europe.
b. Je voudrais dire que j'estime indispensable que les États membres puissent continuer d'autoriser les régions et les communes à percevoir des taxes et que ce domaine ne soit pas uniformément réglé par l'Europe.

[1] DWL models aim at improving the general word choice in the target language (Mauser et al., 2009).

[2] Specifically, 446 instances of pronouns aligned to random words were corrected by hand.

| French | it | | they | |
|---|---|---|---|---|
| | # | % | # | % |
| ça | 79 | 0.43 | 1 | 0.02 |
| cela | 585 | 3.19 | 22 | 0.33 |
| elle | 2,392 | 13.03 | 93 | 1.40 |
| il | 5,332 | 29.04 | 275 | 4.14 |
| ce | 1,919 | 10.45 | 128 | 1.93 |
| elles | 101 | 0.55 | 911 | 13.72 |
| ils | 158 | 0.86 | 3,263 | 49.13 |
| on | 360 | 1.96 | 97 | 1.46 |
| OTHER | 7,432 | 40.48 | 1,852 | 27.88 |
| Total | 18,358 | 100.00 | 6,642 | 100.00 |

Table 1: Distribution of the French translations of English pronouns *it* and *they* in the training data described in Section 4.1.

(5) a. We have that opportunity right now. Let us grasp **it**.
b. Cette chance se présente aujourd'hui, et nous devons **la** saisir !

Examples (2) to (5) are taken from the Europarl section of the data. Table 1 also shows why the problem of pronoun translation is hard: there is no 1-1 correspondence between any English and French pronoun.

Moreover, even if only pronoun-to-pronoun translations are considered, there is no equal distribution of the genders in French. Because all impersonal uses of the pronoun *it* are translated into French *il*, the balance of learning algorithms such as language models is often tilted in favor of the masculine translation. Something similar happens with *they*. In principle, this pronoun can be translated either as *ils* or *elles*; nevertheless, all the members of the group it refers to must be feminine in order to use the feminine *elles*, making this translation much rarer.

## 4 Cross-lingual Pronoun Prediction

### 4.1 Data and Tools

Both sides of the parallel data provided for the shared task are parsed using the Fips parser (Wehrli, 2007). This is a rule-based parser which produces an information-rich phrase-structure representation with predicate-argument labels. Besides, it can also be used as a tagger, generating a POS-tag (containing disambiguated morphological information) and a grammatical function for each word of a given sentence. We relied on this tagger output for extracting most of our features. An example of the output is given in Figure 1.

| | | | |
|---|---|---|---|
| And | CONJ-COO | and | |
| it | PRO-PER-3-SIN | it | SU |
| 's | VERB-IND-PRE-3-SIN | be | |
| a | DET-SIN-NEU | a | FO |
| very | ADV-INT | very | |
| easy | ADJ | easy | |
| question | NOUN-SIN-NEU | question | |
| . | PUNC-POINT | | |

Figure 1: Example of the tagger output of the Fips parser for the sentence *"And it's a very easy question"*. The first column contains the words in the sentence, the second the POS-tags and morphological analysis, the third consists of the lemmas and the fourth of the predicate-argument labels.

For the French side, a unique placeholder is inserted in the place of each REPLACE_XX. This ensures coherent syntactic analysis by the parser, since projections are based on the lexical properties of the heads. The placeholder was inserted in the lexicon as a token with all possible morphological features: both masculine and feminine gender, singular and plural number and the three possible persons. Due to its rule-based nature, the parser unifies only the compatible feature values on each sentence. Consequently, the placeholder allowed us to retrieve some information from the unification process with the verb.

The final training data consists of 25,000 examples composed from a subset of the shared-task data. It includes 747 instances from the TED talks, 14,561 from News Commentary and 9,691 from EuroParl. All systems are built using the Stanford Maximum Entropy package (Manning and Klein, 2003).

### 4.2 Features

We use three types of features roughly following the categorization of Friedrich and Palmer (2014). Most of them rely on the predicate-argument structure of the English side and morphological analysis of the French side. The rationale for this choice is to simulate an MT scenario (where target sentences are not available) in which one could parse the source language to find the argument of interest and may use a dictionary for getting the target-language correspondent morphology. The possible values of all features are listed in Table 2. For each training example, we

extracted the following information:

**Syntactic Features** These features refer to the arguments present in the English sentence (fourth column in Figure 1). Once an argument is identified in the English sentence, the gender and number of the word-aligned French token (most often the head) is retrieved. In the case of the sentential objects, only the values YES or NO are assigned.[3]

1. Current sentence subject
2. Current sentence object
3. Current sentence predicative object
4. Current sentence sentential object
5. Previous sentence subject
6. Previous sentence object
7. Previous sentence predicative object
8. Previous sentence sentential object

**Morphological Features** This information concerns the POS and morphological tags (second column in Figure 1) of the words in the immediate context of each pronoun to predict.

9. Gender and number of all adjectives
10. Previous word POS-tag
11. Following word POS-tag
12. Voice of following verb
13. Person and number of following verb

To obtain the value for feature 9, all adjectives in the previous and the current sentence are identified and the gender and number of their French word-aligned token is searched. Then French gender and number information is aggregated and the most frequent one is selected.

**Context Features** This last set of features refers to the preceding or following tokens of each French pronoun to predict. For these, sentence boundaries are ignored. If the previous word happened to be the full stop of the previous sentence, a full stop is then taken as the value for previous word token.

14. Previous lemma
15. Following lemma
16. Previous word token
17. Following word token
18. Second following word token

### 4.3 System 1

Features 1 and 5 refer to subjects, which are likely to be pronouns aligned with REPLACE_XX items on the French side. In order to simulate the use of an unmodified parser, we dropped the morphological features obtained by unification for the REPLACE_XX items and inserted the special feature value PRON instead. Table 3 contains the obtained results.

[3] Sentential objects are sentences acting as complements of the verb and very often with a conjunction or preposition as their head; therefore, we did not look for gender and number.

| Features | Values |
|---|---|
| 1,2,3,5,6,7,9 | {SIN-FEM, SIN-MAS, PLU-FEM, PLU-MAS, INN-FEM, INN-MAS} |
| 4,8 | {YES, NO} |
| 10,11 | {NOUN,VERB,ADV,PRO,CONJ, PUNC,DET,ADJ,PREP} |
| 12 | {ACTIVE, PASSIVE} |
| 13 | {1-SIN, 1-PLU, 2-SIN, 2-PLU, 3-SIN, 3-PLU} |
| 14,15 | e.g. {*le,avoir,venir,être,rester,...*} |
| 16,17,18 | e.g. {*la,ont,viennent,sont,restent,...*} |

Table 2: Possible values for each of the features. INN stands for *unknown number*.

### 4.4 System 2

For this second experiment, we use the unified values for REPLACE_XX subjects (features 1 and 5). Additionally, the vast OTHER class was split in two classes in order to reduce the imbalance: i) translations by a pronoun not considered among the classes or by a lexical NP, and ii) translations without any pronoun in French. The labels for the latter were taken from the annotation furnished with the training data. After classification, the two subclasses were merged again. The obtained results are presented in Table 3.

### 4.5 Discussion

From the results of System 1 and System 2, it can be noted that the absence of syntactic features (columns **M+C** in Table 3) seems to have a rather small impact in the final results. The syntactic features are motivated in the salience hierarchies established within linguistic theories of salience and AR. In these theories, a syntactically salient argument such as the subject, is more likely to be the antecedent of a pronoun. Our results show, however, that this particular set of features does not contribute much knowledge to the model, and in some cases it only adds noise, as shown by an increase in the scores of columns **M+C**.

Morphology features, on their part, influence the pronouns with feminine and masculine forms, i.e. *il, elle, ils, elles*. However, results are ambiguous: for System 1 there is a positive effect, but for

| | System 1 | | | | System 2 | | | |
|---|---|---|---|---|---|---|---|---|
| **Prediction** | **S+M** | **S+C** | **M+C** | **S+M+C** | **S+M** | **S+C** | **M+C** | **S+M+C** |
| ce | 21.19 | 61.02 | 62.03 | 61.06 | 23.20 | 65.95 | 62.43 | 64.66 |
| cela | 0 | 17.91 | 9.68 | 14.71 | 0 | 20.59* | 9.38 | 19.67 |
| elle | 14.68 | 33.55 | 36.73 | 35.29 | 25.40 | 38.93* | 36.92 | 36.48 |
| elles | 33.33 | 27.03 | 20.25 | 31.33 | 32.50 | 36.11* | 20.25 | 32.10 |
| il | 29.51 | 44.22 | 37.91 | 44.23 | 27.13 | 50.19 | 38.22 | 47.52 |
| ils | 70.34 | 70.80 | 75.07 | 75.88 | 68.97 | 69.16 | 75.00 | 76.13* |
| on | 0 | 32.35 | 24.62 | 30.99 | 10.42 | 31.58 | 26.23 | 34.00 |
| ça | 0 | 5.66 | 5.61 | 9.17 | 0 | 9.35 | 5.61 | 7.48 |
| OTHER | 72.60 | 74.73 | 76.45 | 75.87 | 71.61 | 73.09 | 76.29 | 75.69 |

Table 3: Comparison of F1 scores (%) obtained in the test set with different groups of features. F1 scores were computed using the shared-task scorer. S+M+C correspond to results submitted to the shared-task. *Best results throughout all the systems presented here.

System 2 there is a negative effect columns **S+C** in Table 3). Pronoun *on* is affected in the same way, although we observed that many occurrences referred to a passive construction in English such as (6).

(6) a. *en* ..., if they're given the right work
b. *fr*...si l'on leur confie la bonne mission.

Systems 1 and 2 additionally show that context features are highly important. When they are removed from the model (columns **S+M**), an important drop in the performance is observed. They are particularly determinant for the *ça* and *cela* classes. We had the hypothesis that these pronouns were determined instead by sentential objects, either from the current or the previous sentence.

Looking at the features individually (Table 4), it can be noted that for both systems the morphology information of the following verb (feature 13) is the most important parameter, which makes sense since the task deals mostly with subject pronouns. The other top-ranking features are the following word POS-tag (18), the following lemma (17) and the previous predicative object (10).

The hierarchy in Table 4 reveals further understanding about the context features as well. Features concerning lemmas (15 and 14) have almost as much weight as features concerning raw tokens (16, 17, 18), especially the following lemma. Their influence depends on the pronoun to predict: while raw tokens are determining for pronouns *ce, ça* and *on*, lemmas are determining for pronouns *il, elle, ils* and *elles*.

Furthermore, as depicted in Figure 2, results

| System | Feature number |
|---|---|
| **System 1** | 13,18,10,17,15,16,14,12,11, 8,2,3,5,9,6,7,1,4 |
| **System 2** | 13,18,17,10,15,16,14,11,12 1,4,8,2,9,5,3,6,7 |

Table 4: Features of the model ordered from the most to the least informative.

from System 2 are better[4] than those of System 1 for all the classes. This evidences misclassification due to the big OTHER class, in particular of the less frequent classes. Our two-way distinction is straightforward using the provided data, but we suspect that a finer distinction could further improve results. One could for instance use parsing to distinguish between subject pronouns and object pronouns (such as examples (4), (5)).

The distance between a pronoun and its antecedent is implicitly handled by a language model within a limited window when computing n-gram probabilities. In an attempt to model the notion of distance between the pronoun and each of the arguments in the sentence, we did some tests with the position of each argument as a feature (these were numerical features, then treated as real values). This did not change anything to the model, therefore we dropped it early on.

### 4.6 System 2b and System 2c

Knowing that the test set is composed of TED data, we build an in-domain classifier, System 2b, using only the TED and IWSLT14 corpus for training. Otherwise, it is identical to System 2 (i.e.

[4] $\tau$ = -12.1579, df = 1104, p-value < 2.2e-16

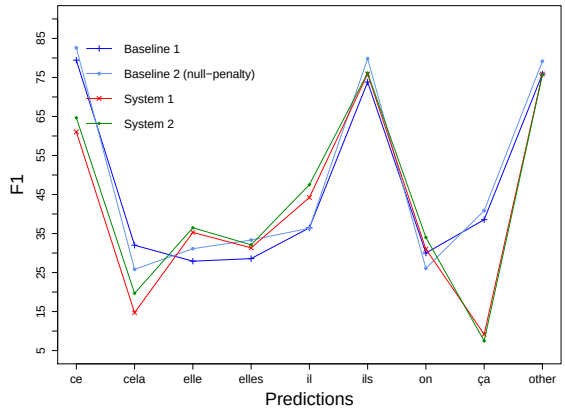


Figure 2: Comparison of fine-grained F-scores of the submitted systems and the task baselines.

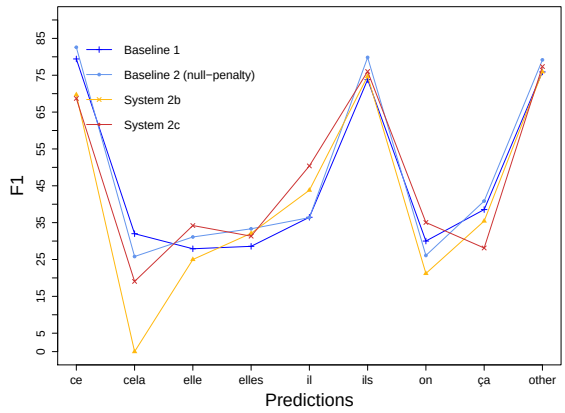


Figure 3: Comparison of fine-grained F-scores of System 2b, System 2c and the task baselines.

with splitting of the OTHER class). Since the training data is much smaller, with only 6,543 training examples, we expect results to be lower than in previous experiments. Results are presented in Table 5.

Results for this system show that context features benefit from the similarity between training and testing data. However, this is not true for pronouns which are determined morphologically as shown by the results of System 1.

Last, the initial training data (25,000 examples from TED, News Commentary and EuroParl) is combined with the 5,796 examples from the IWSLT14 corpus for building System 2c. Results are presented in Table 5. In comparison to System 1 or System 2, the additional training data improves the classification of pronouns *ça* and *ce* (due to the same-domain effect), and additionally,

| | Features S+M+C | |
|---|---|---|
| **Prediction** | **System 2b** | **System 2c** |
| ce | 69.71* | 68.70 |
| cela | 0 | 19.05 |
| elle | 25.00 | 34.21 |
| elles | 32.10 | 31.33 |
| il | 43.80 | 50.39* |
| ils | 74.71 | 76.02 |
| on | 21.21 | 35.05* |
| ça | 35.43* | 28.12 |
| OTHER | 75.93 | 77.36* |

Table 5: Comparison of F1 scores (%) obtained in the test set using the shared-task scorer. *Best results throughout all the systems presented here.

it has a small improvement on the pronouns *il* and *on*. Figure 3 presents a comparison of these two systems with the shared-task baselines.

## 5 Conclusions and Future Work

The selection of features in our experiments showed that the role of syntax is rather small in determining the translation of the English pronouns *it* and *they*. Morphological features on the other hand, had an effect on the prediction of gender-determined pronouns, i.e. feminine and masculine in the case of French. However, we think that more experiments are necessary in order to fully exploit their potential, for instance, with languages with more than two genders. Last, context features proved to be of particular importance to all the classes, above all when the training and testing data are similar. This stress the relevance of the language model for the translation of pronouns and explains the high performance of the baseline as well.

Moreover, our experiments show undoubtedly that splitting the OTHER class improves performance. We think that this a clear step to take in our future work.

Finally, we think that if the notion of *animacy* could be formalized and used as feature, some of the classes would benefit. For instance, it could help to distinguish between human or non-human antecedents, a determining factor for distinguishing between and *it* translated either as *ce* or *il/elle* (Moore et al., 2013). In all the cases, there is plenty of room for improvement.

## References


Robert De Beaugrande and Wolfgang Dressler. 1981. *Introduction to Text Linguistics*. Longman Linguistics Library, Essex.

Annemarie Friedrich and Alexis Palmer. 2014. Automatic prediction of aspectual class of verbs in context. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 517–523, Baltimore, Maryland. Association for Computational Linguistics.

Liane Guillou. 2012. Improving pronoun translation for statistical machine translation. In *Proceedings of the Student Research Workshop at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 1–10, Avignon, France, April. Association for Computational Linguistics.

Christian Hardmeier and Marcello Federico. 2010. Modelling pronominal anaphora in statistical machine translation. In *Proceedings of the 7th International Workshop on Spoken Language Translation*, pages 283–289.

Christian Hardmeier, Jörg Tiedemann, and Joakim Nivre. 2013. Latent anaphora resolution for cross-lingual pronoun prediction. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, EMNLP, pages 380–391, Seattle, Washington. Association for Computational Linguistics.

Christian Hardmeier, Preslav Nakov, Sara Stymne, Jörg Tiedemann, Yannick Versley, and Mauro Cettolo. 2015. Pronoun-focused MT and cross-lingual pronoun prediction: Findings of the 2015 DiscoMT shared task on pronoun translation. In *Proceedings of the Second Workshop on Discourse in Machine Translation*, DiscoMT 2015, Lisbon, Portugal.

Christian Hardmeier. 2014. *Discourse in Statistical Machine Translation*. Ph.D. thesis, Department of Linguistics and Philology, Uppsala University.

Ronan Le Nagard and Philipp Koehn. 2010. Aiding pronoun translation with Co-Reference Resolution. In *Proceedings of the Joint 5th Workshop on Statistical Machine Translation*, pages 258–267, Uppsala, Sweden.

Christopher Manning and Dan Klein. 2003. Optimization, MaxEnt models, and conditional estimation without magic. In *Tutorial at HLT-NAACL and 41st ACL conferences*, Edmonton, Canada and Sapporo, Japan.

Arne Mauser, Saša Hasan, and Hermann Ney. 2009. Extending statistical machine translation with discriminative and trigger-based lexicon models. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1 - Volume 1*, EMNLP'09, pages 210–218, Stroudsburg, PA. Association for Computational Linguistics.

Ruslan Mitkov and Catalina Barbu. 2002. Using bilingual corpora to improve pronoun resolution. *Languages in Contrast*, 4(2):201–211.

Ruslan Mitkov. 2001. Outstanding issues in anaphora resolution. In Alexander Gelbukh, editor, *Computational Linguistics and Intelligent Text Processing*, volume 2004, pages 110–125. Springer Berlin Heidelberg.

Ruslan Mitkov. 2002. *Anaphora Resolution*. Pearson Education Limited, Harlow.

Joshua Moore, Christopher J.C. Burges, Erin Renshaw, and Wen-tau Yih. 2013. Animacy detection with voting models. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 55–60, Seattle, Washington, USA, October. Association for Computational Linguistics.

Vincent Ng. 2010. Supervised noun phrase coreference research: The first fifteen years. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1396–1411.

Michal Novák, Anna Nedoluzhko, and Zdeněk Žabokrtský. 2013. Translation of "It" in a deep syntax framework. In *Proceedings of the Workshop on Discourse in Machine Translation*, pages 51–59, Sofia, Bulgaria. Association for Computational Linguistics.

Michal Novák. 2011. Utilization of anaphora in machine translation. In *Proceedings of the 20th Annual Conference of Doctoral Students–Contributed Papers: Part I*, WDS11, pages 155 — 160, Prague. Matfyzpress.

Andrei Popescu-Belis, Thomas Meyer, Jeevanthi Liyanapathirana, Bruno Cartoni, and Sandrine Zufferey. 2012. Discourse-level annotation over europarl for machine translation: Connectives and pronouns. In *Proceedings of the 8th International Conference on Language Resources and Evaluation*, LREC'12, Istanbul, Turkey. European Language Resources Association (ELRA).

Manfred Stede. 2012. *Disourse Processing*. Morgan and Claypool Publishers, Toronto.

Veselin Stoyanov, Claire Cardie, Nathan Gilbert, Ellen Riloff, David Buttler, and David Hysom. 2009. Conundrums in noun phrase coreference resolution: Making sense of the state-of-the-art. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing*, ACL-IJCNLP 2009.

Michael Strube. 2007. Corpus-based and machine learning approaches to coreference resolution. In Monika Schwarz-Friesel, Manfred Consten, and Mareile Knees, editors, *Anaphors in Text. Cognitive, Formal and Applied Approaches to Anaphoric Reference*, pages 207–222. John Benjamins Publishing Company, Amsterdam.

Eric Wehrli. 2007. Fips, a "Deep" linguistic multilingual parser. In *Proceedings of the Workshop on Deep Linguistic Processing*, pages 120–127. Association for Computational Linguistics.

Jochen Stefan Weiner. 2014. Pronominal anaphora in machine translation. Master of science, Karlsruhe Institute of Technology.