It-disambiguation and source-aware language models for cross-lingual pronoun prediction

It-disambiguation and source-aware language models for cross-lingual pronoun prediction

Sharid Loáiciga
Département de Linguistique
University of Geneva
sharid.loaiciga@unige.ch

Liane Guillou
CIS
LMU Munich
liane@cis.uni-muenchen.de

Christian Hardmeier
Dept. of Linguistics & Philology
Uppsala University
christian.hardmeier@lingfil.uu.se

Abstract

We present our systems for the WMT 2016 shared task on cross-lingual pronoun prediction. The main contribution is a classifier used to determine whether an instance of the ambiguous English pronoun "it" functions as an anaphoric, pleonastic or event reference pronoun. For the English-to-French task the classifier is incorporated in an extended baseline, which takes the form of a source-aware language model. An implementation of the source-aware language model is also provided for each of the remaining language pairs.

1 Introduction

The WMT 2016 shared task on cross-lingual pronoun prediction focuses on the translation of the subject position pronouns "it" and "they" for several language pairs (Guillou et al., 2016). Both of these pronouns perform multiple functions in text, and disambiguation is required if they are to be translated correctly into other languages (Guillou, 2016). The pronoun "they" is typically used as an anaphoric pronoun, but may also be used generically, for example in "*They* say it always rains in Scotland". The pronoun "it" may be used as an anaphoric, pleonastic or event reference pronoun. Examples of these pronoun functions are provided in Figure 1.

anaphoric	I have a **bicycle**. **It** is red.
pleonastic	**It** is raining.
event	He lost his job. **It** came as a total surprise.

Figure 1: Examples of different pronoun functions

Anaphoric pronouns corefer with a noun phrase (i.e. the *antecedent*). *Pleonastic* pronouns, in contrast, do not refer to anything but are required to fill the subject position in many languages, including English, French and German. *Event reference* pronouns may refer to a verb, verb phrase, clause or even an entire sentence.

Different French pronouns are required when translating an instance of "it" depending on its function. For example, anaphoric "it" may be translated with the third-person singular pronouns "il" [masc.] and "elle" [fem.], or with an non-gendered demonstrative such as "cela". The French pronoun "ce" may function as both an event reference and a pleonastic pronoun, but "il" is used only as a pleonastic pronoun.

As revealed in an analysis of the systems submitted to the DiscoMT 2015 shared task on pronoun translation (Hardmeier et al., 2015a), the translation of pleonastic and event reference pronouns poses a particular problem for MT systems (Guillou and Hardmeier, 2016). Poor performance may be attributed to the inability of the systems to disambiguate the various possible functions of the pronoun "it". In the case of systems that incorporate coreference resolution and methods for identifying instances of pleonastic "it", inaccurate output may harm translation performance. No suitable tools exist for the detection of event reference pronouns in English.

To address the problem of disambiguating the function of "it", we propose a classifier that uses information from the current and previous sentences, as well as external tools, and indicates for each instance of "it" whether the pronoun function is anaphoric, pleonastic or event reference. The classifier was trained using data from the ParCor corpus (Guillou et al., 2014) and the *DiscoMT2015.test* dataset (Hardmeier et al., 2016). In both corpora, pronouns are labelled according to their function, following the ParCor annotation scheme. The classifier is incorporated

Proceedings of the First Conference on Machine Translation, Volume 2: Shared Task Papers, pages 581–588,
Berlin, Germany, August 11-12, 2016.

in an extended baseline system for the English-to-French task. The extended baseline takes the form of a n-gram language model that operates over target-language lemmas, but also has access to the identity of the source-language pronouns. Source-aware language models are also provided for the other tasks: English-to-German, German-to-English and French-to-English.

2 Previous Work

Work on pronoun translation, in which a complete machine translation pipeline is provided, has also considered different functions of the pronoun "it". Le Nagard and Koehn (2010) identify and exclude instances of pleonastic "it" in their English-to-French system. Guillou (2015) distinguishes between anaphoric vs. non-anaphoric pronouns in an English-to-French automatic post-editing system. Novák et al. (2013) consider the translation of three different uses of "it" in English-to-Czech translation: *referential it*, referring to a noun phrase, *anaphoric it*, referring to a verb phrase, and *pleonastic it*. These three categories correspond to those that we refer to as anaphoric, event reference and pleonastic, respectively.

Work by Navarretta (2004) and Dipper et al. (2011) has focused on resolving abstract anaphora in Danish and on the manual annotation of abstract anaphora in English and German. Abstract anaphora, in which pronouns refer to abstract entities such as facts or events, is referred to as event reference in this paper. The automatic detection of instances of pleonastic "it" has been addressed by NADA (Bergsma and Yarowsky, 2011), and also by the Stanford sieve-based coreference resolution system (Lee et al., 2011).

The cross-lingual pronoun prediction task formalised by Hardmeier (2014) was first introduced as a shared task at DiscoMT 2015 (Hardmeier et al., 2015a). The participants used a range of features in their classifiers, but this paper marks the first attempt to incorporate a component to disambiguate the various uses of "it".

3 Disambiguating "it"

3.1 Data

The ParCor corpus (Guillou et al., 2014) and *DiscoMT2015.test* dataset (Hardmeier et al., 2016) were used to train the classifier. Under the ParCor annotation scheme, which was used to annotate both corpora, pronouns are labelled according to their function. For all instances of "it" labelled as anaphoric, pleonastic or event reference, the sentence-internal position of the pronoun and the sentence itself are extracted[1]. The pronouns "this" and "that", when used as event reference pronouns, may in many cases be used interchangeably with the pronoun "it" (Guillou, 2016). Consider Ex. 1, in which the pronouns "this" and "it" may be used to express the same meaning.

(1) John arrived late. [This/it] annoyed Mary.

To increase the number of training examples, instances of event reference "this" and "that" are replaced with "it" and added to the training data.

The data was divided into 1504 instances for training, and 501 each for the development and test sets. All sentences were shuffled before the corpus was divided, promoting a balanced distribution of the classes (Table 1).

Data Set	*it-* Event	Anaphoric	Pleonastic	Total
Training	504	779	221	1504
Dev	157	252	92	501
Test	169	270	62	501
Total	830	1301	375	2506

Table 1: Distribution of classes in the training data

All classifiers were trained using the Stanford Maximum Entropy package (Manning and Klein, 2003).

3.2 Features

To parse the corpus, we used the joint part-of-speech tagger and dependency parser of Bohnet et al. (2013) from the Mate toolkit. We used the pre-trained models for English that are available online[2]. In addition, the corpus was lemmatised using the TreeTagger lemmatiser (Schmid, 1994). Although other tools were used, we relied on the output of these two parsers to extract most of our features.

For each training example, we extract the following information:

1. Previous three tokens. This includes words and punctuation. It also includes the tokens in the previous sentence when the *it-* occupies the first position of the current sentence.

[1] A small number of instances of "it" are labelled as cataphoric or extra-textual in the corpora. These are excluded from the classifier training data.

[2] https://code.google.com/p/mate-tools/downloads/list

2. Next two tokens

3. Lemmas of the next two tokens

4. Head word. As the task is limited to subject *it* and *they*, most of the time the head word is a verb.

5. Whether the head word takes a 'that' complement (verbs only)

6. Tense of head word (verbs only). This is computed using the rules described in Loáiciga et al. (2014).

7. Presence of 'that' complement in previous sentence. A binary feature which follows Navarretta (2004)'s conclusion (for Danish) that a particular demonstrative pronoun (*dette*) is often used to refer to the last mentioned situation in the previous sentence, often expressed in a subordinated clause.

8. Predications head. This refers to the predicative complements of the verbs *be*, *appear*, *seem*, *look*, *sound*, *smell*, *taste*, *feel*, *become* and *get*.

9. Closest noun phrase (head) to the left

10. Closest noun phrase (head) to the right

11. Presence of a cleft construction. A binary feature which refers to constructions containing adjectives which trigger extraposed sentential subjects as in '*So it's difficult to attack malaria from inside malarious societies, [...].*

12. Closest adjective to the right

13. VerbNet selectional restrictions of the verb. VerbNet (Kipper et al., 2008) specifies 36 types of argument that verbs can take. We limited ourselves to the values of 'abstract', 'concrete' and 'unknown'.

14. Lemma of the head word

15. Likelihood of head word taking an event subject (verbs only). An estimate of the likelihood of a verb taking a event subject was computed over the Annotated English Gigaword v.5 corpus (Napoles et al., 2012). We considered two cases where an event subject appears often and may be identified by exploiting the parse annotation of the Gigaword corpus. The first case is when the subject is a gerund and the second case is composed of "this" pronoun subjects.

16. NADA probability. The probability that the non-referential "it" detector, NADA (Bergsma and Yarowsky, 2011), assigns to the instance of "it".

We also experimented with other features and options. For features 2 and 3, a window of three tokens showed a degradation in performance. For features 9 and 10, we experimented with adding their WordNet type (WordNet (Princeton University, 2010) contains 26 types of nouns), but this had no effect. The feature combination of noun and adjectives to the left or right also had no effect.

3.3 Results

For development and comparison we built two different baselines. One is a 3-gram language model built using KenLM (Heafield, 2011) and trained over a modified version of the annotated corpus in which every *it* is concatenated with its type (e.g. *it_event*). For testing, the *it* position is filled with each of the three *it_label* and the language model is queried. This baseline functions in a very similar way to the share-task own baseline.

Table 2 presents the results of this baseline using 14-fold cross-validation and a single held-out test set (all test-set mentions refer to the same test set). The motivation for the choice of the number of folds is threefold. First, we wanted to respect document boundaries; second, we aimed for a fair proportion of the three classes in all folds; and, lastly, we tried to lessen the variance given the relatively small size of the corpus. The second baseline is a setting in which all instances of the test set are set to the majority class *it-anaphoric*.

A quick scan of Tables 2 and 3 anticipates one of the conclusions of this paper: predicting event reference pronouns is a complex problem. The 3-gram baseline appears to be biased towards the pleonastic class, as suggested by its high precision and very low recall for the event and anaphoric classes and the opposite situation for the pleonastic class. While our own classifier is more balanced, it achieves only moderate results with the event class. Compared to both of the baselines, it shows only a very small improvement.

	14-fold cross-validation		
	Precision	Recall	F1
it- anaphoric	0.5985	0.2475	0.3502
it- pleonastic	0.1521	0.6213	0.2444
it- event	0.5275	0.2772	0.3633
	Test-set		
	Precision	Recall	F1
it- anaphoric	0.7320	0.2629	0.3869
it- pleonastic	0.1387	0.6935	0.2312
it- event	0.5213	0.2899	0.3726
	Test-set majority class		
	Precision	Recall	F1
it- anaphoric	0.5389	1	0.7004

Table 2: Baselines for the classification of the three types of *it*.

A manual inspection of the results shows that discriminating between anaphoric and event reference instances of *it* is indeed a very subtle process. Determining the presence or the lack of a specific (np-like) antecedent requires the understanding of the complete coreference chain. Take for instance the following example taken from a dialogue in the corpus:

> $_1$You're part of a generation that grew up with the Internet, and it seems as if you become offended at almost a visceral level when you see something done that you think will harm the Internet. $_2$Is there some truth to it? $_3$*It* is. $_4$I think it's very true. $_5$This is not a left or right issue. $_6$Our basic freedoms, and when I say our, I don't just mean Americans, I mean people around the world, *it*'s not a partisan issue .

In the example above the first italicised *it* is an event reference pronoun while the second is an anaphoric pronoun. With access to the whole coreference chain, one can see that the *it* in sentence 3 refers to the event expressed in the first sentence, therefore it is annotated as an event. This same entity is then referred to with the word *issue* in sentence 5, which in turn becomes the antecedent to the *it* in sentence 6. The classifier, however, labelled these two instances as anaphoric and event respectively.

It is worth noting that from the 2031 segments composing the annotated corpus, 349 (17%) contain co-occurrences of between 2 and 7 *it* pronouns within the same segment. We experimented including the previous *it-label*, when there are several within the same sentence, as an additional feature and obtained important gains in performance. It can be seen in the *w/ oracle feature* section of Table 3 that performance improves in almost all cases when this feature is used. The only exception is for the *it-pleonastic* class of the test set. We then tried to approximate this feature by using the relative position of the *it-label* to other *it-labels* within the same sentence (e.g., first, second, etc.). Contrary to the oracle feature, the approximated feature did not lead to any improvement. Modelling co-occurrences of pronouns seems like a promising step in future work.

Binary classification (event vs. non-event) consistently underperformed when compared to the three class set-up.

4 Source-Aware Language Model

The pronoun prediction part of our models is based on an n-gram model over target lemmas similar to the official shared task baseline. In addition to the pure target lemma context, our model also has access to the identity of the source language pronoun, which, in the absence of number inflection on the target words, provides valuable information about the number marking of the pronouns in the source and opens a way to inject the output of the pronoun type classifier into the system.

Our source-aware language model is an n-gram model trained on an artificial corpus generated from the target lemmas of the parallel training data (Figure 2). Before every REPLACE tag occurring in the data, we insert the source pronoun aligned to the tag (without lowercasing or any other processing). The alignment information attached to the REPLACE tag in the shared task data files is stripped off. In the training data, we instead add the pronoun class to be predicted. Note that all REPLACE tags are placeholders for one word translations guaranteed to correspond to a source pronoun *it* or *they* according to the shared-task data preparation (Hardmeier et al., 2015b; Guillou et al., 2016). The n-gram model used for this component is a 6-gram model with modified Kneser-Ney smoothing (Chen and Goodman, 1998) trained

	Dev				Test			
w/o oracle feature	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
it- anaphoric	0.703	0.685	0.758	0.719	0.707	0.716	0.756	0.735
it- pleonastic	0.884	0.758	0.543	0.633	0.936	0.750	0.726	0.738
it- event	0.715	0.545	0.541	0.543	0.703	0.564	0.521	0.542
w/ oracle feature	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
it- anaphoric	0.725	0.705	0.778	0.740	0.727	0.729	0.785	0.756
it- pleonastic	0.886	0.746	0.576	0.650	0.926	0.705	0.694	0.699
it- event	0.739	0.586	0.567	0.576	0.729	0.611	0.538	0.572

Table 3: Classification results of the three types of *it* on the development and test sets.

Source:	**It** 's got these fishing lures on the bottom .
Target lemmas:	**REPLACE_0** avoir ce leurre de pêche au-dessous .
Solution:	*ils*
LM training data:	It **REPLACE_*ils*** avoir ce leurre de pêche au-dessous .
LM test data:	It **REPLACE** avoir ce leurre de pêche au-dessous .

Figure 2: Data for the source-aware language model

with the KenLM toolkit (Heafield, 2011).

To predict classes for an unseen test set, we first convert it to a format matching that of the training data, but with a uniform, unannotated REPLACE tag used for all classes. We then recover the tag annotated with the correct solution using the `disambig` tool of the SRILM language modelling toolkit (Stolcke et al., 2011). This tool runs the Viterbi algorithm to select the most probable mapping of each token from among a set of possible alternatives. The map used for this task trivially maps all tokens to themselves with the exception of the REPLACE tags, which are mapped to the set of annotated REPLACE tags found in the training data.

The source-aware language model described here is identical to the language model component included in the UU-Hardmeier submission (Hardmeier, 2016).

5 English-French "it" Disambiguation System

We used the classifier described in Section 3 to annotate all instances of *it* from the source side of the data which were mapped to a REPLACE item according to the alignment provided. Afterwards, a new source-aware language model is trained in the manner described in Section 4. In this way, instead of the sentence *'It 's got these fishing lures on the bottom .'* presented in Figure 2, the system receives the labelled input *'It_anaphoric 's got these fishing lures on the bottom .'* All the data provided for the shared-task was used in training this system.

6 Results and Analysis

Unfortunately, following the submission of our system we identified an error related to the feature extraction process. We relied on contextual information of the previous sentence for some of our features. However, due to the $1 : N$ alignments, the context information was sometimes inaccurate. The correction of this problem produced the results reported in the section titled *Submitted corrected* in Table 4. The macro-averaged recall obtained is 57.03%, which is considerably better than the result of the submitted system (48.92%), but still slightly lower than the score of 59.84% which was obtained by the unmodified system.

However, some pronouns present better scores using the *submitted corrected* system than the unmodified system. Precision, in particular, is higher (bolded scores in Table 4). This outcome is expected for the pronoun *cela*, which is the French neuter demonstrative pronoun frequently used for event reference. However, there are also gains in precision for *on*, *elles* and *ils*. In our opinion, this suggests that while not directly treating any of the other source-language pronouns (in the context of this shared-task, other source pronouns refers only to *they*), the disambiguation of *it* positively affects the translation of the other target-language

Submitted - w/o labels R: 59.84%

Pronoun	Precision	Recall	F1
ce	89.66	76.47	82.54
elle	40.00	60.87	48.28
elles	27.27	12.00	16.67
il	63.24	70.49	66.67
ils	67.82	83.10	74.68
cela	76.47	41.94	54.17
on	36.36	44.44	40.00
OTHER	88.37	89.41	88.89

Submitted - w/ labels R: 48.92%

Pronoun	Precision	Recall	F1
ce	70.11	89.71	78.71
elle	0.00	0.00	0.00
elles	20.00	16.00	17.78
il	70.97	36.07	47.83
ils	50.96	74.65	60.57
cela	48.65	58.06	52.94
on	42.86	33.33	37.50
OTHER	86.59	83.53	85.03

Submitted corrected - w/ labels R: 57.03%

Pronoun	Precision	Recall	F1
ce	89.09	72.06	79.67
elle	31.25	43.48	36.36
elles	**30.77**	16.00	21.05
il	54.43	70.49	61.43
ils	**69.41**	83.10	75.64
cela	**86.67**	41.94	56.52
on	**40.00**	44.44	42.11
OTHER	85.71	84.71	85.21

Table 4: Final system

pronouns. The pronoun *it*, after all, is used three times more frequently than *they* in the training data (Loáiciga and Wehrli, 2015).

Looking at the predictions, we confirmed that both source-aware language models produced identical results almost all of the time, with the system without the labels producing more correct predictions in total. However, there are some few examples where the system with the labels outperforms both the baseline and the un-labelled one. A contrastive example can be seen in Figure 3.

7 Conclusions and Future Work

Distinguishing between anaphoric and event reference realisations of "it" is a very complex task. In particular, it can be difficult to determine the antecedent of an event reference pronoun. The identification of pleonastic realisations, on the other hand, is almost impossible in an n-gram context such as that provided by a language model. However, it is feasible in the three class setting, and at the same time helpful for the disambiguation of the event and anaphoric realisations.

Source:	**it_anaphoric** just takes a picture of objective reality as **it_anaphoric** is .
LM w/o labels:	il OTHER
LM w/labels:	elle OTHER
Baseline:	cela OTHER
Gold	**elle** prendre juste un image objectif de **la** réalité .

Figure 3: Examples of predictions of the final systems. The Gold translation is lemmatized.

While our results are modest, they point towards an improvement in the general quality of pronoun translation. Accurate disambiguation of the pronoun "it" has the potential to help NLP applications such as Machine Translation and Coreference Resolution.

In the near future, we will experiment with other classification algorithms suitable for small training sets. We also intend to experiment with features that incorporate semantic knowledge in the form of statistics computed over external resources, including the Gigaword corpus. Last, with the generated data from this shared-task, we plan to do bootstrap and experiment with self-training.

Acknowledgments

SL was supported by the Swiss National Science Foundation under grant no. P1GEP1_161877. CH and LG were supported by the Swedish Research Council under project 2012-916 *Discourse-Oriented Statistical Machine Translation*. Large-scale computations were performed on the Abel cluster, owned by the University of Oslo and the Norwegian metacenter for High Performance Computing (NOTUR), under project nn9106k.

References

Shane Bergsma and David Yarowsky. 2011. NADA: A robust system for non-referential pronoun detection. In Iris Hendrickx, Sobha Lalitha Devi, António Branco, and Ruslan Mitkov, editors, *Anaphora Processing and Applications: 8th Discourse Anaphora and Anaphor Resolution Colloquium (DAARC)*, Lecture Notes in Artificial Intelligence, pages 12–23. Springer, Faro, Portugal.

Bernd Bohnet, Joakim Nivre, Igor Boguslavsky, Richárd Farkas, Filip Ginter, and Jan Hajič. 2013. Joint morphological and syntactic analysis for richly inflected languages. *Transactions of the Association for Computational Linguistics*, 1:415–428.

Stanley F. Chen and Joshua Goodman. 1998. An empirical study of smoothing techniques for language modeling. Technical report, Computer Science Group, Harvard University, Cambridge (Mass.).

Stefanie Dipper, Christine Rieger, Melanie Seiss, and Heike Zinsmeister. 2011. Abstract anaphors in german and english. In Iris Hendrickx, Sobha Lalitha Devi, António Branco, and Ruslan Mitkov, editors, *Anaphora Processing and Applications: 8th Discourse Anaphora and Anaphor Resolution Colloquium (DAARC)*, Lecture Notes in Artificial Intelligence, pages 96–107. Springer, Faro, Portugal.

Liane Guillou and Christian Hardmeier. 2016. PROTEST: A test suite for evaluating pronouns in machine translation. In *Proceedings of the Eleventh Language Resources and Evaluation Conference (LREC'16)*, Portorož (Slovenia), May.

Liane Guillou, Christian Hardmeier, Aaron Smith, Jörg Tiedemann, and Bonnie Webber. 2014. ParCor 1.0: A parallel pronoun-coreference corpus to support statistical MT. In *Proceedings of the Tenth Language Resources and Evaluation Conference (LREC'14)*, pages 3191–3198, Reykjavík (Iceland).

Liane Guillou, Christian Hardmeier, Preslav Nakov, Sara Stymne, Jörg Tiedemann, Yannick Versley, Mauro Cettolo, Bonnie Webber, and Andrei Popescu-Belis. 2016. Findings of the 2016 WMT shared task on cross-lingual pronoun prediction. In *Proceedings of the First Conference on Machine Translation (WMT16)*, Berlin, Germany. Association for Computational Linguistics.

Liane Guillou. 2015. Automatic Post-Editing for the DiscoMT Pronoun Translation Task. In *Proceedings of the Second Workshop on Discourse in Machine Translation*, pages 65–71, Lisbon, Portugal. Association for Computational Linguistics.

Liane Guillou. 2016. *Incorporating Pronoun Function into Statistical Machine Translation*. Ph.D. thesis, University of Edinburgh.

Christian Hardmeier, Preslav Nakov, Sara Stymne, Jörg Tiedemann, Yannick Versley, and Mauro Cettolo. 2015a. Pronoun-focused MT and cross-lingual pronoun prediction: Findings of the 2015 DiscoMT shared task on pronoun translation. In *Proceedings of the 2nd Workshop on Discourse in Machine Translation (DiscoMT 2015)*, pages 1–16, Lisbon (Portugal).

Christian Hardmeier, Preslav Nakov, Sara Stymne, Jörg Tiedemann, Yannick Versley, and Mauro Cettolo. 2015b. 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, Jörg Tiedemann, Preslav Nakov, Sara Stymne, and Yannick Versely. 2016. DiscoMT 2015 Shared Task on Pronoun Translation. LINDAT/CLARIN digital library at Institute of Formal and Applied Linguistics, Charles University in Prague. `http://hdl.handle.net/11372/LRT-1611`.

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

Christian Hardmeier. 2016. Pronoun prediction with latent anaphora resolution. In *Proceedings of the First Conference on Machine Translation (WMT)*, Berlin (Germany).

Kenneth Heafield. 2011. KenLM: faster and smaller language model queries. In *Proceedings of the Sixth Workshop on Statistical Machine Translation*, pages 187–197, Edinburgh (Scotland, UK), July. Association for Computational Linguistics.

Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. 2008. A large-scale classification of english verbs. *Language Resources and Evaluation Journal*, 42(1):21–40.

Ronan Le Nagard and Philipp Koehn. 2010. Aiding Pronoun Translation with Co-Reference Resolution. In *Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR*, pages 252–261, Uppsala, Sweden. Association for Computational Linguistics.

Heeyoung Lee, Yves Peirsman, Angel Chang, Nathanael Chambers, Mihai Surdeanu, and Dan Jurafsky. 2011. Stanford's Multi-Pass Sieve Coreference Resolution System at the CoNLL-2011 Shared Task. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*, CONLL Shared Task '11, pages 28–34, Portland, Oregon. Association for Computational Linguistics.

Sharid Loáiciga and Éric Wehrli. 2015. Rule-based pronominal anaphora treatment for machine translation. In *Proceedings of the Second Workshop on Discourse in Machine Translation*, DiscoMT 2015, Lisbon, Portugal.

Sharid Loáiciga, Thomas Meyer, and Andrei Popescu-Belis. 2014. English-french verb phrase alignment in europarl for tense translation modeling. In *Proceedings of the 9th International Conference on Language Resources and Evaluation*, LREC'14, pages 674–681, Reykjavik, Iceland. European Language Resources Association (ELRA).

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.

Courtney Napoles, Matthew Gormley, and Benjamin Van Durme. 2012. Annotated gigaword. In *Proceedings of the Joint Workshop on Automatic Knowledge Base Construction and Web-scale Knowledge Extraction*, AKBC-WEKEX, pages 95–100, Montreal, Canada. Association for Computational Linguistics.

Costanza Navarretta. 2004. Resolving individual and abstract anaphora in texts and dialogues. In *Proceedings of the 20th International Conference on Computational Linguistics*, COLING'04, pages 233–239, Geneva, Switzerland. Association for Computational Linguistics.

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.

Princeton University. 2010. Wordnet.

Helmut Schmid. 1994. Probabilistic part-of-speech tagging using decision trees. In *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK.

Andreas Stolcke, Jing Zheng, Wen Wang, and Victor Abrash. 2011. SRILM at sixteen: Update and outlook. In *Proceedings of the IEEE Automatic Speech Recognition and Understanding Workshop*, Waikoloa (Hawaii, USA).