C O P *i* L CAMBRIDGE OCCASIONAL PAPERS *in* LINGUISTICS
Volume **10**, Article 11: 267–278, 2017 | ISSN 2050-5949

The challenges of cross-linguistic parsing[*]

Maja Gwóźdź
University of Cambridge

1 Introduction

This short paper aims at outlining the present state of cross-linguistic parsing, which, in the context of computational linguistics,[1] can be defined as "the process of automatically analyzing a given sentence, viewed as a sequence of words, in order to determine its possible underlying syntactic structures" (Nederhof & Satta 2010: 105, cf. Naumann & Langer 1994: 3). Parsing is an essential component of machine translation and question answering systems, amongst others (cf. Naumann & Langer 1994: 13). Figure 1 is an example[2] of parsing based on phrase structure rules.

```
(ROOT
(S
 (S
          (VP (VB Read)
            (NP (DT the) (JJ entire) (NN article))))
 (: ;)
 (S
    (NP (EX there))
    (VP (VBZ 's)
    (NP (DT a) (NN punchline))
    (, ,)
    (ADVP (RB too))))
 (. .)))
```

Figure 1 Phrase structure parse of the sentence *Read the entire article; there's a punchline, too.*

Depending on the features of the computational task, parsers are commonly modelled on phrase structure grammars, dependency grammars,[3] categorial grammars, or Tree Adjoining Grammars. They can also be oriented towards the Government

* I would like to thank Dr Andrew Caines for his comments on the earlier drafts.

1 I will use the terms *computational linguistics* and *natural language processing* interchangeably, even though there exist definitions that point out the subtle differences between these fields.

2 This representation was obtained by using the online version of the Stanford parser, available here: `http://nlp.stanford.edu:8080/parser/index.jsp` (last accessed 18th Dec 2016).

3 It is worth mentioning that there exist isomorphisms (which can be extended to the corresponding probabilistic versions) between certain types of dependency and phrase structure grammars (see the theoretical discussion in Manning & Schütze 1999: 429–430).

and Binding Theory (so-called 'parameter-based parsing'). The latter appears to be a less popular approach nowadays but it has been demonstrated, for instance by Fong & Berwick (1992), that this method can be very productive. In natural language processing (NLP) systems, parsing can be perceived as the central component due to the fact that "the accuracy of the parses can have much impact on the success of an application as a whole" (Nederhof & Satta 2010: 105, cf. Merlo, Bunt & Nivre 2010: 11–12). Although parsing is, in itself, an intriguing NLP challenge from a theoretical point of view (see Nederhof & Satta 2010, Jurafsky & Martin 2014: 45–84), its application to a wide range of interdisciplinary problems makes it even more significant. For instance, a recent study by Taboada, Meizoso, Martínez, Riaño, & A. (2011) has shown that parsing can be successfully implemented in clinical tasks.

In this paper, I will address the most common challenges of cross-linguistic parsing (I will assume that cross-linguistic parsing involves different source and target languages) by evaluating the adaptation of English parsers to processing data from morphologically rich languages (henceforth, MRLs), such as Arabic, Hebrew, French, Polish, or Turkish, to name but a few (for recent advances in morphological tagging of MRLs see Acedański 2010, Marsza lek-Kowalewska, Zaretskaya & Souček 2014, Jaafar, Bouzoubaa, Yousfi, Tajmout & Khamar 2016; see Legrand & Collobert 2016 for an innovative technique of parsing MRLs; Seddah, Tsarfaty, Kübler, Candito, Choi, Farkas, Foster, Goenaga, Gojenola, Goldberg, Green, Habash, M., Maier, Nivre, Przepiórkowski, Roth, Seeker, Versley, Vincze, Woliński, A. & de la Clérgerie 2013). More specifically, I will focus on Polish and make generalisations about other MRLs based on the results obtained from processing Polish data with an English-trained parser.

Intuitively, it is clear that parsers trained on language-specific datasets are not likely to perform well in cross-linguistic tasks, especially if the target language is typologically unrelated to the source language. There is sufficient evidence in the literature to support this hypothesis (cf. Arun & Keller 2005: 1, Tsarfaty, Seddah, Kübler & Nivre 2013: 16, Green, de Marneffe & Manning 2013: 196). When discussing the challenge of domain adaptation (both minimally supervised and unsupervised) within the same language, Dell'Orletta, Marchi, Montemagni, Venturi, Agnoloni & Francesconi (2013: 58) (cf. Merlo et al. 2010: 3–4) assert that: "In spite of the fact that nowadays dependency parsing can be carried out with high levels of accuracy, the adaptation of parsers to new domains without target domain training data remains an open issue". An in-depth investigation of parsing and domain-dependency has also been carried out by McClosky (2010). He focused on English constituency parsing and demonstrated the effectiveness of self-training (a semi-supervised method).

The paper is organised as follows. First, I provide an overview of the recent advances in cross-linguistic parsing and summarise the most challenging problems posed by this procedure. In order to demonstrate the ineffective adaptation of English parsing models to MRLs, I have conducted a simple experiment involving the processing of Polish data with English parsers. I have tested the dependency parsing models due to the availability of both models in the same format, which allows for a more accurate comparison. The reason for adopting the dependency-based approach will be clarified in the next section. The relevant technicalities are described in the methodology section. The outcome of the experiment is analysed in the remaining part of the paper and this leads to general conclusions about the present state of cross-linguistic parsing.

2 Cross-linguistic parsing

Perhaps unsurprisingly, English has been one of the most extensively studied languages with respect to parsing. In the 1990s the constituency-based parsing models of English set the "performance ceiling of 92% F_1-score[4] using the PARSEVAL evaluation metrics" (Black, Abney, Flickinger, Gdaniec, Grishman, Harrison, Hindle, Ingria, Jelinek, Klavans, Liberman, Marcus, Roukos, Santorini & Strzalkowski 1991 quoted in Tsarfaty, Seddah, Goldberg, Kübler, Versley, Candito, Foster, Rehbein & Tounsi 2010: 1; footnote added). However, when applied to other languages, these parsers proved significantly less successful. The deterioration in performance was initially attributed to the differences in annotation schemata, as well as the inadequacy of the PARSEVAL evaluation metrics. Later it was observed that the unsatisfactory performance rate had been caused by mere typological differences between English (a language with relatively poor morphology) and MRLs. The following features of MRLs still pose difficulties for English-trained parsers: "a large inventory of word-forms, higher degrees of word order freedom, and the use of morphological information in indicating syntactic relations" Tsarfaty et al. (2010: 1).

The dependency-based format, on the other hand, is widely perceived to be a better methodological and computational (in terms of efficiency) choice for representing (not necessarily parsing) MRLs than constituency-based models (Tsarfaty et al. 2010: 3; cf. the higher accuracy of dependency models in parsing French in Arun & Keller 2005: 311–312, Buchholz & Marsi 2006: 149–150, Merlo et al. 2010: 2–3). This is due to the fact that dependency structures are not bound by the linear word order, as opposed to the constituency-based representations. To clarify this point, the basic premise of dependency grammar is that "syntactic structure essentially consists of words linked by binary, asymmetrical relations called dependency relations" and that a "dependency relation holds between a syntactically subordinate word, called the dependent, and another word on which it depends, called the head" (Kübler, McDonald & Nivre 2009: 2). It is well-known that MRLs are much more flexible with regard to word order than English, therefore dependency structures seem to be better suited for capturing discontinuous constituents (cf. the notions of *nonprojectivity* and *non-configurational information* in Tsarfaty et al. 2010: 2, 6). An example of dependency parsing is presented in 2 (cf. the phrase structure tree for the same sentence above).

[4] In crude terms, the F_1-score may be defined as the harmonic mean of precision and recall computed by means of the following formula: $\frac{2(precision \times recall)}{precision + recall}$.

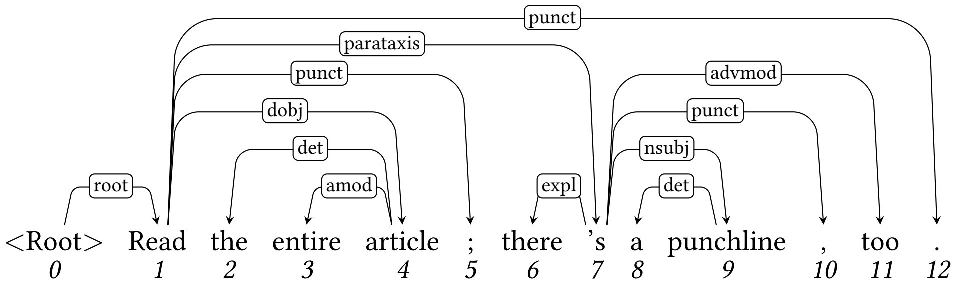

Figure 2 Dependency parsing: An example from the gold standard English set.

Recently, cross-linguistic parser adaptation has been extensively studied, for example, by Smith & Smith (2004), Arun & Keller (2005), or Smith & Eisner (2011). Søgaard (2011: 682) claims that cross-language dependency parser adaptation "is similar to, but more difficult than most domain adaptation or transfer learning scenarios, where differences between source and target distributions are smaller". In the same vein, Smith & Eisner (2011: 823) call this problem "an extreme case of out-of-domain data". The main aims of building cross-language parsers are, amongst others, economy, i.e., maintaining one tool instead of having to develop multiple parametrised models, and more accurate parsing of data containing non-native vocabulary. To cite Ammar, Mulcaire, Ballesteros, Dyer & Smith (2016: 431): "code-switching or code-mixing [...], which is pervasive in some genres, in particular social media, presents a challenge for monolingually-trained NLP models". The most obvious challenge posed by cross-linguistic parsing can be summarised in the question: How can a parser trained on a dataset from one language perform the necessary inductive steps and correctly decode the structure of sentences from a language to which it has never been exposed?

As regards some methods proposed to solve specific problems in cross-linguistic parser adaptation, Arun & Keller (2005) demonstrated that lexicalised models (in this context the term *lexicalisation* refers to removing part-of-speech tags and relying only on lexical material in parsing) yield satisfactory results for both English and French (partly) due to the fact that these languages exhibit a non-flexible word order. An intermediate step between parsing data from the same source and target language and true cross-linguistic parsing is, perhaps, the CONLL-X shared task initiative (Buchholz & Marsi 2006, cf. Ammar et al. 2016), in which participants were given datasets from various languages (as different as Japanese and Turkish, for instance) and had to rely on one parsing model to process the multilingual data. Although in this paper I do not consider it as cross-linguistic parsing proper, multilingual parsing is, undoubtedly, a substantial improvement towards constructing truly universal systems. So far, bilingual parsing appears to be the most fruitful area of research. For instance, Burkett (2012) reports that parsing supported by word alignment of Chinese and English data improves the output quality (cf. Smith & Smith 2004).

3 Methodology

All computational tasks were performed on a MacBook with 1.4 GHz Intel Core i5. For the purposes of this experiment I chose to work with MaltParser[5] due to the availability of both English and Polish[6] dependency models (see Wróblewska 2012 and Wróblewska & Woliński 2012 for the description of the Polish model; cf. Marsza lek-Kowalewska et al. 2014) and its popularity among NLP researchers (cf. Eragani & Kuchibhotla 2014). MaltParser is an example of a "data-driven approach to dependency parsing that has been applied to a range of different languages, consistently giving a dependency accuracy in the range of 80–90%" (Nivre, Hall, Nilsson, Chanev, Eryigit, Kübler, Marinov & Marsi 2007: 95). On a more technical note, the methodology on which MaltParser is built consists primarily of inductive learning and deterministic parsing (consult Nivre et al. 2007 for more information).

After installing MaltParser and obtaining the English and Polish dependency models, I downloaded[7] the datasets in the conll-u format (compatible with MaltParser) for those languages. Both datasets contained training, development, and test files. The fully optimised pre-trained English and Polish dependency models served as a means of reference (regarding dependency relations) and were not used for comparative purposes. This was due to the fact that both models had been designed to conform to the conll format, while the datasets for custom parser training had been prepared in the conll-u format (a version of the conll-x format). To illustrate this format, figure 3 is provided (see below).

Moreover, there were slight differences in the annotation conventions. However, the results below will show that this factor did not undermine the validity of my experiment. Nevertheless, it ought to be mentioned that annotation differences are often cited as a serious problem in parser evaluation because it is impossible to distinguish between parsing errors proper and those caused by incompatible standards (see Tsarfaty et al. 2010: 4, Nivre 2015: 5).

The implementation of MaltOptimizer[8] (a useful tool for automatic MaltParser optimisation) designed specifically for the Shared Task on Parsing Morphologically Rich Languages (stpmrl) seemed especially well-suited for my study. In order to obtain the most accurate results, I initially used MaltOptimizer (Ballesteros & Nivre 2012) for the English data and MaltOptimizer stpmrl for the Polish data. Unfortunately, the data validation task in MaltOptimizer could not be performed

[5] Freely available at: http://www.maltparser.org/download.html. In this paper, I am referring to the latest release (1.9.0, as of 31st Oct 2016).

[6] The English model is freely available at: http://nlp.stanford.edu/software/stanford-dependencies.shtml#Methods (last accessed 31st Oct 2016). See also http://maltparser.org/mco/english_parser/engmalt.html for more information. The Polish model can be downloaded at: http://zil.ipipan.waw.pl/PolishDependencyParser?action=fullsearch&value=linkto%3A%22PolishDependencyParser%22&context=180 (last accessed 31st Oct 2016).

[7] http://stp.lingfil.uu.se/~nivre/download/UD_Data+MaltEval.tar.gz (last accessed 3rd Nov 2016); the treebanks come from the Universal Dependencies project (http://universaldependencies.org; last accessed 4th Nov 2016).

[8] Available here: http://nil.fdi.ucm.es/maltoptimizer/download.html (last accessed 3rd Nov 2016). See also the stpmrl implementation: http://nil.fdi.ucm.es/maltoptimizer/spmrl.html (last accessed 3rd Nov 2016).

1	Read	_	VERB	VB	_	0	root	_	_
2	the	_	DET	DT	_	4	det	_	_
3	entire	_	ADJ	JJ	_	4	amod	_	_
4	article	_	NOUN	NN	_	1	dobj	_	_
5	;	_	PUNCT	,	_	1	punct	_	_
6	there	_	DET	EX	_	7	expl	_	_
7	's	_	VERB	VBZ	_	1	parataxis	_	_
8	a	_	DET	DT	_	9	det	_	_
9	punchline	_	NOUN	NN	_	7	nsubj	_	_
10	,	_	ADV	RB	_	7	advmod	_	_
11	too	_	ADV	RB	_	7	advmod	_	_
12	.	_	PUNCT	.	_	1	punct	_	_

Figure 3 A sentence from the English gold standard represented in the CONLL-U format.

on my computer due to processing issues or a software error. I then decided to train the parsers without any optimising procedures. A similar strategy was applied by Søgaard (2011: 684) in his cross-linguistic parsing study. To my mind, the development of English and Polish parsing models from scratch ensured maximum objectivity. The full justification of this methodological choice will become apparent later on. The comparison of parsing efficiency of the unoptimised and properly parametrised models is presented in the next section.

The learning and parsing steps were performed in accordance with the MaltParser user guide and the PARSEME instructions.[9] I then used the two models to parse the respective test datasets. It ought to be emphasised that, at this stage, the source language was still the same as the target language. The evaluation of the results was performed with the MaltEval tool. It ought to be noted that due to some annotation discrepancies in the English and Polish CONLL-U datasets, only the English model could be successfully evaluated with MaltEval. In order to assess the accuracy of the Polish model manually and compared (line by line) my custom model against the gold standard. The final and, by far, most important stage of the experiment involved parsing Polish data with the English model. In order to compare the types of cross-linguistic mislabelling, I have also used the Polish parsing model to process English data (4.187 s) and observed that the model was, unsurprisingly, equally unsuccessful.

4 Results and discussion

As already signalled in the previous section, MaltEval was used to measure the accuracy of the English model. In order to provide evidence for the incompatibility of the pre–trained optimised English models (differing with respect to the underlying

[9] http://maltparser.org/userguide.html. PARSEME instructions: stp.lingfil.uu.se/~nivre/research/parseme_lab.html (last accessed 3rd Nov 2016).

algorithms) and the CONLL-U datasets, I additionally parsed the test file with the two parametrised models. The results are presented in the table below:

	pre-trained English model with a polynomial kernel	**pre-trained English model based on support vector machines**	**the unoptimised English model**
row mean (accuracy per token)	0.46	0.467	0.828
row count	25165	25165	25165

Table 1 The LAS accuracy of two pre-trained English models and the unoptimised model.

The LAS (Labelled Attachment Score) is a popular metric used for evaluating the performance of dependency parsers. It is computed as: the number of tokens with correct heads and labels divided by the total number of tokens. As can be easily seen, the discrepancy in the level of accuracy between the pre-trained models and the unoptimised model is rather striking. It is also worth noting that the difference between the results obtained by applying the two models, one using support vector machines with a polynomial kernel and the other linear support vector machines, is negligible. This observation is confirmed by the description provided by the models' developers.[10] The scores presented above justify the development of an unoptimised parsing model.

The learning stage of the Polish model took only 70.767 s and the parsing of test data lasted 3.072 s (63.473 s and 4.053 s for the English model, respectively). The evaluation of the Polish model against the gold standard (that is, the original annotated test set) was achieved manually. I compared the two files by examining elements that did not match. The qualitative comparison was performed by searching for the mismatched rows and examining wrongly parsed fragments. The similarity of the files was then measured by means of a simple technique, namely, the `wdiff` command. Both files contained precisely 73446 words and shared 71910 tokens (98%), differing only with respect to 1536 tokens (2%). No deletions or insertions were recorded, thus the discrepancies in the two files stemmed from substitutions. A careful analysis indicated that the substitution patterns were random and that there was no tendency towards misidentifying a particular dependency relation. It ought to be stressed that the 2% difference mentioned above does not describe the accuracy of the parser. Judging by the qualitative and quantitative analyses, it can be safely estimated that the model reached an accuracy in the range of ca 75–80%.

Having trained and evaluated the two models, I then proceeded to the last part of the experiment, that is, I used the unoptimised English model to parse Polish

[10] See http://maltparser.org/mco/english_parser/engmalt.html (last accessed 6th Nov 2016).

data (2.077 s). The procedure was analogous to the steps described above and all settings were set to default. As expected, the results were very unsatisfactory. Even though the statistics obtained by evoking the `wdiff` command seemed to be fairly optimistic, they had to be interpreted with great care. It turned out that the files shared as many as 60413 tokens (82%), there were no deletions/insertions, and 'only' 13033 words (18%) differed. Superficially, this score is extremely satisfactory. However, a quick glance at the actual content of the files provided enough evidence to state that the parser trained on English data misidentified every dependency relation and assigned the label `root` to it. As expected, these findings have shown that an unoptimised English model without any contact with the target language in the training phase cannot perform well in processing MRLs. As a reminder, the accuracy of the English model when the source and target language was the same was as high as 82.8%. When applied to a different target language it was 0%.

It is interesting to note that the labels assigned by the Polish model, albeit incorrect in almost all cases, were much more varied than the default `root` label overused by the English model. Quantitatively speaking, the Polish model preferred the `amod` label (adjectival modifier) and sometimes, due to chance, this prediction was compatible with the English gold standard. Quite confusingly, the Polish model misidentified punctuation marks and instead of the obvious `punct` label assigned a complement dependency to it. These results demonstrate that applying the simplest (and most desirable) method of cross-linguistic parsing is not yet possible if we expect even a remotely decent output. However, performing more cross-linguistic experiments whose outcome is known to be a failure might also lead to valuable observations. The comparison of mislabelling, especially from a qualitative perspective, could be one starting point towards fine-tuning cross-linguistic parsers. Other factors which might have affected the overall score were the minor differences in annotation standards and the size of the two sets:

Polish training set	**English training set**	**Polish test set**	**English test set**
739398	2045860	73446	251650

Table 2 Comparison of the size of training and test sets for Polish and English.

The English training and test sets are considerably larger than the corresponding Polish datasets (approximately 36% and 29% difference, respectively). To my mind, these discrepancies are negligible because even if we were to standardise the sets, the results would not change drastically. It can be safely speculated that the English model would still assign the default `root` label to every line and the Polish model would probably exhibit the same predilection towards the `amod` label (some variation in the assignment of other labels would probably be observed).

5 Conclusions

This short essay has shown that the challenges posed by true cross-linguistic parsing are numerous and ought to be solved separately. At present, multilingual and word-alignment-based bilingual parsing seem to be the most promising areas of research which could contribute to a more thorough understanding of parsing data from two morphologically dissimilar languages. The simple experiment described in this paper has confirmed the intuition regarding the ineffective adaptation of English models to parsing MRLs data. By training two dependency parsing models on the datasets from English and Polish (an MRL), which shared a common format, I obtained two high-performing parsers with respective accuracy rates of 82.8% and ca 75–80%. However, the application of the English model to the parsing of Polish data (and vice versa) revealed a complete misidentification of dependency relations. The mislabelling was entirely predictable but the type of parsing errors differed and could not be easily foreseen. To my mind, it is possible that a more substantial analysis of cross-linguistic parsing errors could reveal useful patterns and facilitate optimisation. The study was meant as a very basic contribution to the challenging and fascinating task of cross-language parsing. This experiment had also certain limitations, such as considerable differences in the size of respective datasets and using unoptimised parsers for the analyses. At present it seems that a satisfactory score in cross-linguistic parsing is still unattainable. Further inquiries into this issue will, undoubtedly, lead to the development of state-of-the-art parsers with a great potential for extreme out-of-domain adaptation.

References

Acedański, S. 2010. A morphosyntactic Brill tagger for inflectional languages. In H. Loftsson, E. Rögnvaldsson & S. Helgadottir (eds.), *Advances in Natural Language Processing: 7th International Conference on NLP, IceTAL 2010, Reykjavik, Iceland, August 16–18, 2010, Proceedings*, 3–14. Berlin/Heidelberg: Springer.

Ammar, W., G. Mulcaire, M. Ballesteros, C. Dyer & N. A. Smith. 2016. Many languages, one parser. In D. Chiang (ed.), *Transactions of the Association for Computational Linguistics*, 431–444. Association for Computational Linguistics. https://arxiv.org/pdf/1602.01595. 7th Nov. 2016.

Arun, A. & F. Keller. 2005. Lexicalization in crosslinguistic probabilistic parsing: The case of French. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, 306–313. Ann Arbor, MI.

Ballesteros, M. & J. Nivre. 2012. MaltOptimizer: An optimization tool for MaltParser. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, 58–62. Avignon, France: Association for Computational Linguistics (ACL). www.aclweb.org/anthology/E12-2012. 7th Nov. 2016.

Black, E., S. Abney, D. Flickinger, C. Gdaniec, R. Grishman, P. Harrison, D. Hindle, R. Ingria, F. Jelinek, J. Klavans, M. Liberman, M. Marcus, S. Roukos, B. Santorini & T. Strzalkowski. 1991. A procedure for quantitatively comparing the syntactic coverage of English grammars. In *Proceedings of the Speech and Natural Language Workshop, Defence Advanced Research Projects Agency (DARPA)*, 306–311. Pacific Grove, Calif.: Morgan Kaufmann.

Buchholz, S. & E. Marsi. 2006. CONLL–X shared task on multilingual dependency parsing. In *Proceedings of the Tenth Conference on Computational Language Learning(CONLL)*, 149–164. New York, NY.

Burkett, D. 2012. Syntactic agreement in bilingual corpora. BA, London: School of Oriental and African Studies, University of London.

Dell'Orletta, F., S. Marchi, S. Montemagni, G. Venturi, T. Agnoloni & E. Francesconi. 2013. Domain adaptation for dependency parsing at Evalita 2011. In B. Magnini, F. Cutugno, M. Falcone & E. Pianta (eds.), *Evaluation of Natural Language and Speech Tools for Italian: International Workshop, EVALITA 2011, Rome, January 24–25, 2012, revised selected papers*, 58–69. Berlin/Heidelberg: Springer.

Eragani, A. K. & V. Kuchibhotla. 2014. Improving Malt Dependency Parser using a simple grammar-driven unlexicalised dependency parser. In *2014 International Conference Asian Language Processing (IALP)*, 211–214. http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6973482. 7th Nov. 2016.

Fong, S. & R. C. Berwick. 1992. Isolating cross-linguistic parsing complexity with a principles-and-parameters parser: A case study of Japanese and English. In *COLING '92: Proceedings of the 14th conference on computational linguistics*, vol. 2, 631–637. Nantes, France.

Green, S., M.-C. de Marneffe & C. D. Manning. 2013. Parsing models for identifying multiword expressions. *Computational Linguistics* 39(1). 195–227. doi:doi:10.1162/ coli_a_00139.

Jaafar, Y., K. Bouzoubaa, A. Yousfi, R. Tajmout & H. Khamar. 2016. Improving Arabic morphological analyzers benchmark. *International Journal of Speech Technology* 19(2). 259–267. doi:10.1007/s10772-016-9340-x.

Jurafsky, D. & J. H. Martin. 2014. *Speech and Language Processing*. Second. Edinburgh: Pearson.

Kübler, S., R. McDonald & J. Nivre. 2009. *Dependency Parsing*. San Francisco, CA: Morgan & Claypool 2nd edn.

Legrand, J. & R. Collobert. 2016. Deep neural networks for syntactic parsing of morphologically rich languages. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (volume 2: short papers)*, vol. 2, 573–578. Association for Computational Linguistics (ACL). https://www.aclweb.org/anthology/P/P16/P16-2093.pdf. 27th Oct. 2016.

Manning, C. & H. Schütze. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

Marsza lek-Kowalewska, K., A. Zaretskaya & M. Souček. 2014. Stanford typed dependencies: Slavic languages application. In A. Przepiórkowski & M. Ogrodniczuk (eds.), *Advances in Natural Language Processing: 9th International Conference on NLP PolTAL 2014, Warsaw, Poland, September 17–19, 2014*, 151–163. Berlin/Heidelberg: Springer.

McClosky, D. 2010. Any domain parsing: Automatic domain adaptation for natural language parsing. Ms.

Merlo, P., H. Bunt & J. Nivre. 2010. Current trends in parsing technology. In H. Bunt, P. Merlo & J. Nivre (eds.), *Trends in Parsing Technology: Dependency Parsing, Domain Adaptation, and Deep Parsing*, 1–18. Dordrecht/Heidelberg/London/New York: Springer.

Naumann, S. & H. Langer. 1994. *Parsing: Eine Einführung in die maschinelle Analyse natürlicher Sprache*. Stuttgart: Teubner.

Nederhof, M.-J. & G. Satta. 2010. Theory of parsing. In A. Clark, C. Fox & S. Lappin (eds.), *The Handbook of Computational Linguistics and Natural Language Processing*, Oxford: Blackwell.

Nivre, J. 2015. Towards a universal grammar for natural language processing. In A. Gelbukh (ed.), *Computational Linguistics and Intelligent Text Processing*, 3–16. Cham: Springer.

Nivre, J., J. Hall, J. Nilsson, A. Chanev, G. Eryigit, S. Kübler, S. Marinov & E. Marsi. 2007. Maltparser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering* 13(2). 95–135. doi:10.1017/S1351324906004505.

Seddah, D., R. Tsarfaty, S. Kübler, M. Candito, J. D. Choi, R. Farkas, J. Foster, I. Goenaga, K. Gojenola, Y. Goldberg, S. Green, N. Habash, Kuhlmann M., W. Maier, J. Nivre, A. Przepiórkowski, R. Roth, W. Seeker, Y. Versley, V. Vincze, M. Woliński, Wróblewska A. & E. V. de la Clérgerie. 2013. Overview of the SPMRL 2013 Shared Task: Cross-framework evaluation of parsing morphologically rich languages. In A. Przepiórkowski & M. Ogrodniczuk (eds.), *Proceedings of the Fourth Workshop on Statistical Parsing of Morphologically Rich Languages*, 146–182. Seattle, Washington, USA: Springer. http://www.aclweb.org/anthology/W13-4917.pdf. 27th Oct. 2016.

Smith, D. A. & J. Eisner. 2011. Parser adaptation and projection with quasi—synchronous grammar features. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, vol. 2, 822–831. http://www.aclweb.org/anthology/D09-1086. 27th Oct. 2016.

Smith, D. A. & N. A. Smith. 2004. Bilingual parsing with factored estimation: Using English to parse Korean. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 49–56. Barcelona, Spain. http://www.ccs.neu.edu/home/dasmith/biparse.pdf. 27th Oct. 2016.

Søgaard, A. 2011. Data point selection for cross-anguage adaptation of dependency parsers. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers*, 686–686. Portland, Oregon. http://www.aclweb.org/anthology/P11-2120. 27th Oct. 2016.

Taboada, M., M. Meizoso, D. Martínez, D. Riaño, & Alonso A. 2011. Combining open-source natural language processing tools to parse clinical practice guidelines. *Expert Systems* 30(1). 3–11. doi:10.1111/j.1468-0394.2010.00575.x.

Tsarfaty, R., D. Seddah, Y. Goldberg, S. Kübler, Y. Versley, M. Candito, J. Foster, I. Rehbein & L. Tounsi. 2010. Statistical parsing of morphologically rich languages (spmrl) what, how and whither. In *Proceedings of the NAACL HLT 2010: First workshop on statistical parsing of morphologically-rich languages*, 1–12.

Tsarfaty, R., D. Seddah, S. Kübler & J. Nivre. 2013. Parsing morphologically rich languages: Introduction to the special issue. *Computational Linguistics* 39(1). 15–22. doi:10.1162/ coli_a_00133.

Wróblewska, A. 2012. Polish dependency bank. *Linguistic Issues in Language Technology* 7(2). 1–18.

Wróblewska, A. & M. Woliński. 2012. Preliminary experiments in Polish dependency parsing. In P. Bouvry, M. A. Klopotek, F. Leprevost, M. Marciniak, A. Mykowiecka & H. Rybinski (eds.), *Security and Intelligent Information Systems: International Joint Confererence 2011 Revised Selected Papers*, 279–292. Berlin/Heidelberg: Springer. doi:10.1007/978-3-642-25261-7_22.

Maja Gwóźdź
University of Cambridge
mkg33@cam.ac.uk