Improving Polish Mention Detection with Valency Dictionary

Maciej Ogrodniczuk
Institute of Computer Science
Polish Academy of Sciences
Jana Kazimierza 5
01-248 Warsaw, Poland
`maciej.ogrodniczuk@ipipan.waw.pl`

Bartłomiej Nitoń
Institute of Computer Science
Polish Academy of Sciences
Jana Kazimierza 5
01-248 Warsaw, Poland
`bartek.niton@gmail.com`

Abstract

This paper presents results of an experiment integrating information from valency dictionary of Polish into a mention detection system. Two types of information is acquired: positions of syntactic schemata for nominal and verbal constructs and secondary prepositions present in schemata. The syntactic schemata are used to prevent (for verbal realizations) or encourage (for nominal groups) constructing mentions from phrases filling multiple schema positions, the secondary prepositions – to filter out artificial mentions created from their nominal components. Mention detection is evaluated against the manual annotation of the Polish Coreference Corpus in two settings: taking into account only mention heads or exact borders.

1 Introduction

Coreference resolution systems are believed to suffer from lack of integration of "deeper" knowledge, with respect to both semantics and world knowledge, while it has been recognized from the very beginning (Hobbs, 1978) that they make very important and at the same time difficult factors in the process[1] and that present attempts of integration of such features are bringing only small improvements to the overall accuracy (see next section for examples). The slow progress in solving complex semantics- or knowledge-related issues

we are experiencing today is promoting the switch into the search of new algorithms and models and probably also adds to the general loss of global interest in coreference resolution.

Nevertheless we argue that such situation should not be considered as failure of semantic approaches but rather as a consequence of enormous dimensions and complication of the knowledge system which needs to be applied to linguistic processing, including reference decoding. On the contrary, we believe that the method of small steps towards the big goal is constantly bringing useful models and resources to the field, year by year growing in size and complexity. It is particularly important for languages other than English where more subtle properties of semantic constructs can influence the results.

In the current paper we show how integration of a relatively simple rule taking into consideration verbal and nominal valency in Polish slightly but consequently improves mention detection scores.

2 Related Work

(Kehler et al., 2004) integrated preferences inferred from statistics of subject–verb, verb–object and possessive–noun predicate–argument frequencies into a pronoun-based resolution system which resulted in 1% accuracy improvement. Several works integrating semantic processing into coreference resolution were also proposed, e.g. (Ponzetto and Strube, 2006b) integrated predicate-argument pairs into (Soon et al., 2001)'s resolution system which yielded 1.5 MUC F_1 score improvement on ACE 2003 data.

(Ponzetto and Strube, 2006a; Ponzetto and Strube, 2007) used Wikipedia, WordNet and semantic role tagging to compute semantic relatedness between anaphor and antecedent to achieve 2.7 points MUC F_1 score improvement on ACE

---

[1]Cf. a concluding sentence from (Sapena et al., 2013): "Although it is clearly necessary to incorporate world knowledge to move forward in the field of coreference resolution, the process required to introduce such information in a constructive way has not yet been found." See also e.g. Michael Strube's presentation at CORBON 2016: "The (Non)Utility of Semantics for Coreference Resolution".

2003 data.

(Rahman and Ng, 2011) labelled nominal phrases with FrameNet semantic roles achieving 0.5 points B^3 and CEAF F_1 score improvement and used YAGO type and means relations achieving 0.7 to 2.8 points improvement on OntoNotes-2 and ACE 2004/2005 data.

(Durrett and Klein, 2013) incorporated in their system shallow semantics by using WordNet hypernymy and synonymy, number and gender data for nominals and propers, named entity types and latent clusters computer from English Gigaword corpus, reaching 1.6 points improvement on gold data and 0.36 points on system data.

For Polish, WordNet and Wikipedia-related features were used to improve verification of semantic compatibility for common nouns and named entities in BARTEK-3 coreference resolution system (Ogrodniczuk et al., 2015, Section 12.3) resulting in improvement of approx. 0.5 points MUC F_1 score. Experiments with integration of external vocabulary resources coming from websites registering the newest linguistic trends in Polish, fresh loan words and neologisms not yet covered by traditional dictionaries have been also performed showing low coverage of new constructs in evaluation data (Ogrodniczuk, 2013).

All these results showed challenges regarding knowledge-based resources, mainly concerning the memory and time complexity of the task as well as low coverage of complex features in the test data, but at the same time brought some (sometimes tiny) improvements to coreference resolution scores.

3   Problem Definition

In our approach *mentions* are defined as text fragments (nominal groups including attached prepositional phrases and relative clauses) which could potentially create references to discourse world objects. Such definition has both syntactic and semantic grounds: inclusion of extensive syntactically dependent phrases into mention borders is important due to semantic understanding of mentions: *pierwszy człowiek na Księżycu* 'the first man on the Moon' or *samochód, który potrącił moją żonę* 'the car which hit my wife' have different meanings than just *człowiek* 'the man' or *samochód* 'the car'. One of the consequences of this distinction is treating as mentions all embedded phrases with heads distinct from the head of

the main phrase (meaning that they corresponded to different entities). Therefore, in the example:

(1)   *szef działu firmy*
      'the head of the branch of the company'

three noun phrases should be considered as mentions referring to, accordingly, 'the head of the branch of the company', 'the branch of the company' and 'the company' itself.

The need of exact mention border detection stands in contradiction with unavailability of a constituency parser for Polish with sufficient coverage[2] which could solve most of the attachment problems. Current state-of-the-art mention detector for Polish (see Section 4.3) identifies nominal groups with a relatively old Spejd shallow parser. Our work attempts to use valency schemata from a recently created valency dictionary for Polish (see Section 4.1) for two purposes: to prevent mention borders to cross positions of a syntactic schema and to filter out mentions created from nominal components of secondary prepositions, also present in the valency dictionary.

4   Resources and Tools

4.1   Walenty, a Polish Valence Dictionary

Walenty (Przepiórkowski et al., 2014)[3] is a comprehensive human- and machine-readable dictionary of Polish valency information for verbs, nouns, adjectives and adverbs. It consists of two interconnected layers, syntactic and semantic, and features precise linguistic description, including the structural case, clausal subjects, complex prepositions, comparative constructions, control and raising and semantically defined phrase types. Lexicon entries have strictly defined formal structure and the represented syntactic and semantic phenomena are always attested in linguistic reality, with the National Corpus of Polish (Przepiórkowski et al., 2012, later referred to as NKJP) as a primary source of data and Internet and linguistic literature as secondary sources.

Each lexical entry is identified by its lemma and consists of a number of syntactic valence schemata with each schema being a set of syntactic positions. Apart from the two labeled argument

---

[2]Currently available constituency parsers for Polish such as Świgra (`http://zil.ipipan.waw.pl/Świgra`) or POLFIE (`http://zil.ipipan.waw.pl/LFG`) do not yet guarantee sufficient coverage of natural language constructs.

[3]See also `http://walenty.ipipan.waw.pl/`.

Figure 1: A sample schema in Walenty.

positions, subject and object, usual phrase types are considered, such as nominal phrases (NP), prepositional phrases (PREPNP), adjectival phrases (ADJP), clausal phrases (CP), etc. Phrase types can be further parameterised by corresponding grammatical categories, e.g., NP and ADJP are parameterised by information concerning case. The underscore symbol '_' denotes any value of a grammatical category, e.g., INFP(_) denotes infinitival phrase of any aspect.

Figure 1 presents a sample schema for the verb *łączyć* ('to link') with subject, object, nominal phrase in the instrumental case and prepositional phrase using preposition *z* ('with') and nominal component in the instrumental case again, as in the following example:

(2)   *Potężne [komputery]* SUBJ *[łączą]* VERB *[firmę]* OBJ *[światłowodami]* NP(INST) *[z cyfrowym światem]* PREPNP(Z,INST)·

'Powerful [computers] SUBJ [link] VERB [the company] OBJ [with the digital world] PREPNP(Z,INST) [using optical fiber] NP(INST)·'

As of January 2017, Walenty contains over 65K schemata for 12K Polish verbs and 16K schemata for about 2500 nouns and is still expanding.

In our experiments we use Walenty in textual format (Hajnicz et al., 2015) which can be downloaded directly from Slowal Web application[4] (Nitoń et al., 2016). The version used in our experiment dates January 17, 2017.

4.2 Polish Coreference Corpus

The Polish Coreference Corpus[5] (Ogrodniczuk et al., 2015) is a large corpus of Polish general nominal coreference built upon NKJP. Each text of the corpus is a 250–350-word sample consisting of full subsequent paragraphs extracted from a larger text. With its 1900 documents from 14 text genres,

containing about 540K tokens, 180K mentions and 128K coreference clusters, the PCC is among the largest manually annotated coreference corpora in the international community.

Mentions in PCC are understood as broadly as possible, with the following components included in the nominal phrase:

1. adjectives adjusting their form (case, number, gender) to the superordinate noun, e.g. *kolorowe kwiaty, duży czerwony tramwaj, lebiodka pospolita* 'colourful flowers', 'big red tram', 'oregano'

2. adjectives in genitive, singular, neuter, e.g. *coś fantastycznego, nic dziwnego* 'something fantastic', 'nothing strange'

3. nouns adjusting its case and number to the superordinate noun (apposition), e.g. *malarz pejzażysta, miasto Łódź* 'a landscape painter', 'the city of Łódź'

4. nouns in genitive, e.g. *kolega brata, protokół przesłuchania* 'a friend of my brother', 'the protocol of the hearing'

5. numeral phrases as subordinate elements of the nominal element, e.g. *zabójca pięciu kobiet* 'the killer of five women'

6. adjective participles adjusting its form to the superordinate noun, together with its subordinate element, e.g. *nadchodzące zmiany, rozbudowany hotel, zapaleńcy prowadzący swoje wojenki* 'oncoming changes', 'expanded hotel', 'hotheads waging their little wars'

7. relative clauses, e.g. *dziewczyna, o której rozmawiamy* 'the girl we talked about'

8. prepositional-nominal phrases, e.g. *ustawa o podatku dochodowym, droga na skróty* 'the law on income tax', 'a way across the country'

9. particles, e.g. *prawie cała rodzina, tylko ty* 'almost the whole family', 'only you'

10. adverbs as adjectives and participle modifiers, e.g. *szalenie ciekawy film* 'incredibly interesting film'.

---

[4]http://zil.ipipan.waw.pl/Slowal
[5]http://zil.ipipan.waw.pl/PCC

Similarly some phrases with syntactic head other than nominal were also considered mentions, such as numeral phrases or coordinated nominal phrases.

The current version of PCC data is 0.92 dated December 29, 2014.

4.3 Mention Detector for Polish

The state-of-the-art mention detection tool for Polish is MentionDetector[6] which uses information from morphosyntactic, shallow syntactic and named entity annotations created with state-of-the-art tools for Polish. MentionDetector is mostly a rule based tool with a statistical mechanism for detecting zero subjects. The following constructs are recognized:

1. single-segment nouns and nominal groups, detected with Spejd shallow parser[7] (Przepiórkowski and Buczyński, 2007) fitted with an adaptation of the NKJP grammar of Polish (Ogrodniczuk et al., 2014)

2. pronouns, identified with a disambiguating morphosyntactic tagger Pantera[8] (Acedański, 2010) with a morphological analyser and lemmatizer Morfeusz[9] (Woliński, 2014)

3. zero subjects, detected with a custom solution (Kopeć, 2014)

4. nominal named entities, detected with Nerf[10] (Waszczuk et al., 2013).

The current version of MentionDetector is 1.3 dated October 13, 2016.

5 The Experiment

The idea for the experiment is based on the observation that delimitation of mentions based on their semantic understanding is different for nominal and verbal constructs: for nominal phrases engaged in valency schemata (making the mention 'core') all syntactic positions should be included into the mention boundaries since they add vital supporting information to the core while for

verbal phrases their nominal or prepositional positions correspond to different semantic roles and cannot be linked into a single mention. This assumption is verified with schemata acquired from Walenty against the PCC gold annotation.

The entry point for both nominal and verbal parts of the experiment is the same: finite verb forms as well as nominal and prepositional phrases are detected in the text[11] and matched against valency schemata. This is achieved by comparing base forms of syntactic heads of words to entries from the valency dictionary (directly for the main Walenty entry and by creating textual representations of phrase types for syntactic positions).

5.1 Nominal realizations

If a nominal schema with two positions corresponding to phrase types detected in the document is found, both the core nominal phrase and the dependent phrases are merged into a single mention, as in:

(3) *Od tamtego czasu miał miejsce [konflikt]* NOUN *[polskiego ambasadora]* NP(GEN) *[z polskim księdzem]* PREPNP(Z,INST).

'Since then there was [a conflict] NOUN [of the Polish ambassador] NP(GEN) [with the Polish priest] PREPNP(Z,INST).'

PREPNP constructions are created from the preposition word (tagged as PREP by Spejd) and the case of the head word from prepositional-nominal groups. NP constructions are created using the case of the nominal group head word.

The results of mention detection after adding this rule to base MentionDetector are presented in Table 1 under *Mention merging*.

5.2 Verbal realizations

If a verbal schema with nominal or prepositional positions is detected in the document, we prevent creation of a single mention out of phrases from different syntactic positions, cf.

(4) [Gratuluję] VERB [Włochom] NP(DAT) [awansu] NP(GEN).

'I [congratulate] VERB [the Italians] NP(DAT) [on their promotion] NP(GEN).'

---

[11]In order to process prepositional phrases Spejd shallow grammar was adapted to detect prepositional-nominal groups (PREPNG).

The results of mention detection after adding this rule are presented in Table 1 under *Mention cleaning*, note that *nominal realizations* rule is also active.

5.3 Secondary Prepositions and Phraseological Compounds

Another valuable information present in Walenty is a list of approx. 200 secondary prepositions used in syntactic schemata[12]. Since secondary prepositions are lexicalized combinations of primary (monomorphemic) prepositions and nominal or prepositional phrases, their nominal components can be often automatically (and always incorrectly) marked as mentions. Table 1 under *Walenty list* presents the results of removal of such mentions from the system set.

The next step was expansion of the list of complex prepositions using other available sources, the first of them being *The PWN Universal Dictionary of the Polish Language*[13] (Dubisz, 2006). Secondly, rules responsible for building secondary prepositions out of individual prepositions and nouns in Spejd grammar were examined and their components were also excluded from the list of mention candidates. Last but not least, Spejd grammar rules for idiomatic expressions (marked as *frazeo*) were investigated to collect indeclinable phraseologic phrases with nominal component (underlined below) such as:

- particle-adverbs (`Qub`), e.g. *bez wątpienia* 'without a doubt'

- adverbs (`Adv`), e.g. *w lot* 'immediately'

- interjections (`Interj`), e.g. *broń Boże* 'heaven forbid'

- adjectives (`Adj`), e.g. *na poziomie* 'ambitious'

- conjunctions (`Conj`), e.g. *przy czym* 'at the same time'

- compounds (`Comp`), e.g. *w miarę jak (słuchali)* 'as (they listened)'

That means that sometimes complex prepositions text strings are not always used as a preposition and we must know the wider text context to distinguish whether they are truly complex prepositions or constructions bringing up mention into the discourse. Spade helps us in this distinction.

The results of mention detection after adding this rule are presented in Table 1 under *Secondary prepositions*, note that *nominal realizations* and *verbal realizations* rules are also active.

6 Results

Results of mention detection follow the procedure described in (Ogrodniczuk et al., 2015). Precision, recall and F-measure are calculated using Scoreference application from the Polish Coreference Toolset[14]. As compared to SemEval approach (Recasens et al., 2010) where systems were rewarded with 1 point for correct mentions boundaries, 0.5 points for boundaries within the gold NP including its head, 0 otherwise, in our evaluation we decided not to reward partial matches but to provide two alternative mention detection scores: EXACT boundary match and HEAD match.

Table 1 compares the results of exact mention detection to the best available mention detection results for Polish. The baseline for our verification is the newest result of evaluation of current version of MentionDetector on PCC test data[15].

Nominal realizations rule increases mention detection by over 1%. We believe that it could be increased even higher with larger dictionary. Our rule is using noun constraints only and by far there are only about 2500 nouns in Walenty. Fortunately Walenty is still expanding and further score improvement is a matter of time.

Verbal realizations rule is bringing very small mention detection score improvement, on the other hand it is highly precise.

Head only detection results are presented for comparison, as we can see they have slightly increased after using *secondary prepositions and phraseological compounds* rule. This is because during this step we have removed a lot of wrong single-segment mentions (consisting of heads only) which has noticeable and positive impact on HEAD mention detection precision.

Configuration	EXACT			HEAD		
	Precision	Recall	F₁	Precision	Recall	F₁
Baseline	67.07%	67.19%	67.13%	88.68%	**89.37%**	89.02%
Mention merging	68.34%	67.95%	68.15%	88.63%	88.74%	88.69%
Mention cleaning	68.35%	**67.96%**	68.16%	88.63%	88.74%	88.69%
Secondary prepositions	**69.59%**	67.85%	**68.71%**	**90.02%**	88.30%	**89.15%**

Table 1: Mention detection evaluation results

7 Conclusions

The presented experiment showed usefulness of valency schemata in the process of mention detection although the scale of improvement was relatively small. It should be attributed to several factors such as the limited size of the valency dictionary or sparsity of cases where valency rules can intervene (as opposed to 'general' cases).

The setting used only two most frequent types of phrases present in valency schemata, nominal and prepositional phrases, so one of the next steps could be analysis how other types of phrases intervene in the process of mention construction.

Even though the gains are far from being huge as compared to the progress introduced to the field in the recent years by adoption of new algorithms and architectures, experiments with integration of knowledge and semantics into the process seem worth pursuing, particularly for languages other than English for which they may offer fine-tuning of the language-independent solutions bringing slow but stable progress to results of linguistic analysis.

Acknowledgements

References

Szymon Acedański. 2010. A morphosyntactic brill tagger for inflectional languages. In Hrafn Loftsson, Eiríkur Rögnvaldsson, and Sigrún Helgadóttir, editors, *Advances in Natural Language Processing: 7th International Conference on NLP, IceTAL 2010, Reykjavik, Iceland, August 16-18, 2010*, pages 3–14, Berlin, Heidelberg. Springer Berlin Heidelberg.

Stanisław Dubisz, editor. 2006. *Uniwersalny słownik języka polskiego PWN*. Wydawnictwo Naukowe PWN. vol. 1–4.

Greg Durrett and Dan Klein. 2013. Easy victories and uphill battles in coreference resolution. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1971–1982, Seattle, Washington, USA, October. Association for Computational Linguistics.

Elżbieta Hajnicz, Bartłomiej Nitoń, Agnieszka Patejuk, Adam Przepiórkowski, and Marcin Woliński. 2015. Internetowy słownik walencyjny języka polskiego oparty na danych korpusowych. *Prace Filologiczne*, LXV:95–110.

Jerry R. Hobbs. 1978. Resolving Pronoun References. *Lingua*, 44:311–338.

Andrew Kehler, Douglas Appelt, Lara Taylor, and Aleksandr Simma. 2004. The (Non)Utility of Predicate-Argument Frequencies for Pronoun Interpretation. In *Proceedings of 2004 North American chapter of the Association for Computational Linguistics annual meeting*, pages 289–296.

Mateusz Kopeć. 2014. Zero subject detection for Polish. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, volume 2: Short Papers*, pages 221–225, Gothenburg, Sweden. Association for Computational Linguistics.

Bartłomiej Nitoń, Tomasz Bartosiak, and Elżbieta Hajnicz. 2016. Accessing and Elaborating *Walenty* — a Valence Dictionary of Polish — via Internet Browser. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*, pages 1352–1359, Portorož, Slovenia. ELRA, European Language Resources Association.

Maciej Ogrodniczuk, Alicja Wójcicka, Katarzyna Głowińska, and Mateusz Kopeć. 2014. Detection of Nested Mentions for Coreference Resolution in Polish. In Maciej Ogrodniczuk and Adam Przepiórkowski, editors, *Advances in Natural Language Processing: Proceedings of the 9th International Conference on NLP, PolTAL 2014*, volume 8686 of *Lecture Notes in Computer Science*, pages 270–277, Warsaw, Poland. Springer International Publishing.

Maciej Ogrodniczuk, Katarzyna Głowińska, Mateusz Kopeć, Agata Savary, and Magdalena Zawisławska. 2015. *Coreference in Polish: Annotation, Resolution and Evaluation*. Walter De Gruyter.

Maciej Ogrodniczuk. 2013. Discovery of common nominal facts for coreference resolution: Proof of concept. In R. Prasath and T. Kathirvalavakumar, editors, *Mining Intelligence and Knowledge Exploration (MIKE 2013)*, volume 8284 of *Lecture Notes in Artificial Intelligence*, pages 709–716. Springer-Verlag, Berlin, Heidelberg.

Simone Paolo Ponzetto and Michael Strube. 2006a. Exploiting Semantic Role Labeling, WordNet and Wikipedia for Coreference Resolution. In *Proceedings of the Human Language Technology Conference of the NAACL, Main Conference*, pages 192–199, New York City, USA, June. Association for Computational Linguistics.

Simone Paolo Ponzetto and Michael Strube. 2006b. Semantic Role Labeling for Coreference Resolution. In *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics: Posters & Demonstrations*, EACL '06, pages 143–146, Stroudsburg, PA, USA. Association for Computational Linguistics.

Simone Paolo Ponzetto and Michael Strube. 2007. Knowledge Derived from Wikipedia for Computing Semantic Relatedness. *Journal of Artificial Intelligence Research*, 30(1):181–212.

Adam Przepiórkowski and Aleksander Buczyński. 2007. Spejd: Shallow Parsing and Disambiguation Engine. In Zygmunt Vetulani, editor, *Proceedings of the 3rd Language & Technology Conference*, pages 340–344, Poznań, Poland.

Adam Przepiórkowski, Mirosław Bańko, Rafał L. Górski, and Barbara Lewandowska-Tomaszczyk, editors. 2012. *Narodowy Korpus Języka Polskiego*. Wydawnictwo Naukowe PWN, Warsaw.

Adam Przepiórkowski, Elżbieta Hajnicz, Agnieszka Patejuk, Marcin Woliński, Filip Skwarski, and Marek Świdziński. 2014. Walenty: Towards a comprehensive valence dictionary of Polish. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*, pages 2785–2792, Reykjavík, Iceland. ELRA.

Altaf Rahman and Vincent Ng. 2011. Coreference Resolution with World Knowledge. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 814–824.

Marta Recasens, Lluís Màrquez, Emili Sapena, M. Antònia Martí, Mariona Taulé, Véronique Hoste, Massimo Poesio, and Yannick Versley. 2010. SemEval-2010 Task 1: Coreference Resolution in Multiple Languages. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 1–8, Stroudsburg, PA, USA. Association for Computational Linguistics.

Emili Sapena, Lluís Padró, and Jordi Turmo. 2013. A Constraint-based Hypergraph Partitioning Approach to Coreference Resolution. *Computational Linguistics*, 39(4):847–884.

Wee Meng Soon, Hwee Tou Ng, and Daniel Chung Yong Lim. 2001. A Machine Learning Approach to Coreference Resolution of Noun Phrases. *Computational Linguistics*, 27(4):521–544.

Jakub Waszczuk, Katarzyna Głowińska, Agata Savary, Adam Przepiórkowski, and Michał Lenart. 2013. Annotation tools for syntax and named entities in the National Corpus of Polish. *International Journal of Data Mining, Modelling and Management*, 5(2):103–122.

Marcin Woliński. 2014. Morfeusz Reloaded. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014)*, pages 1106–1111, Reykjavík, Iceland. European Language Resources Association.