A Hybrid Approach for Deep Machine Translation

Kiril Simov
Linguistic Modelling Department
IICT-BAS
Bulgaria
kivs@bultreebank.org

Petya Osenova
Linguistic Modelling Department
IICT-BAS
Bulgaria
petya@bultreebank.org

Abstract

This paper presents a Hybrid Approach to Deep Machine Translation in the language direction from English to Bulgarian. The set-up uses pre- and post-processing modules as well as two-level transfer. The language resources that have been incorporated are: WordNets for both languages; a valency lexicon for Bulgarian; aligned parallel corpora. The architecture comprises a predominantly statistical component (factor-based SMT in Moses) with some focused rule-based elements. The experiments show promising results and room for further improvements within the MT architecture.

1 Introduction

The paper presents a hybrid approach for Deep Semantic Machine Translation. For that purpose, however, the linguistic phenomena that constitute deep semantics have to be defined. A list of such phenomena have been considered in (Hajič, 2011) and (Bos, 2013), among others. They include but are not limited to the following ones: Semantic Roles (words vs. predicates, Lexical Semantics (Word Sense Disambiguation (WSD)), Multiword Expressions (MWE), Logical Form (LF), Metonymy, Named Entities (NE), Co-reference (pronominal, bridging anaphora), Verb Phrase Ellipsis, Collective/Distributive NPs, Scope (Negation, Quantifiers), Presuppositions, Tense and Aspect, Illocution Force, Textual Entailment, Discourse Structure/ Rhetorical Relations, neo-Davidsonian Events, Background Knowledge, Information Structure etc. All the mentioned phenomena represent various levels of granularity and different linguistic dimensions.

In our deep Machine Translation (MT) system we decided to exploit the following components in the transfer phase: Lexical Semantics (WSD), MultiWord Expression (MWE), Named Entities (NE) and Logical Form (LF). For the incorporation of Lexical Semantics through the exploitation of WordNet and Valency dictionary the knowledge-based approach to WSD has been accepted. Concerning the LF, we rely on Minimal Recursion Semantics (MRS) in its two variants - the full one (MRS) and the more underspecified one (Robust MRS (RMRS)). The MWE and NE are parts of the lexicons. We should note that there are also other appropriate LF frameworks that are briefly mentioned below.

One of the MRS-related semantic formalisms is the Abstract Meaning Representation (AMR[1]), which aims at achieving whole-sentence deep semantics instead of addressing various isolated holders of semantic information (such as, NER, coreferences, temporal anchors, etc.). AMR also builds on the available syntactic trees, thus contributing to the efforts on sembanking. It is English-dependent and it makes an extensive use of PropBank framesets (Kingsbury and Palmer, 2002) and (Palmer et al., 2005). Its concepts are either English words or special keywords. AMR uses approximately 100 relations. They include: frame arguments, general semantic relations, relations for quantities and date-entities, etc.

The Groningen Meaning Bank (GMB) integrates various phenomena in one formalism. It has a linguistically motivated, theoretically solid (CCG[2]/DRT[3]) background.

---

[1]http://www.isi.edu/natural-language/amr/a.pdf
[2]Combinatory Categorical Grammar
[3]Discourse Representation Theory

In this paper the NLP strategies are presented for Hybrid Deep Machine Translation in the direction from English-to-Bulgarian. Under Hybrid MT we understand the usage of the automatic Moses system together with a rule-based component at the transfer phase.

The paper is structured as follows:in section 2 the components of the hybrid MT architecture is presented. Section 3 discusses the deep semantic processing. Section 4 reports on the current experiments and results. Section 5 concludes the paper.

2 A Hybrid MT Architecture

Our Hybrid MT Architecture (see Fig. 1) includes the following components: NLP preprocessing of the source language (in this case English)(statistical approach); projection of the annotations on the tokens in the text (statistical); first-level transfer where the source tokens are substituted as much as possible with their translations in the target language and an intermediate language is created (rule-based and statistical); second-level transfer where the translations are based on the lemma (statistical); generation is performed (statistical) and post-processing is applied (rule-based).
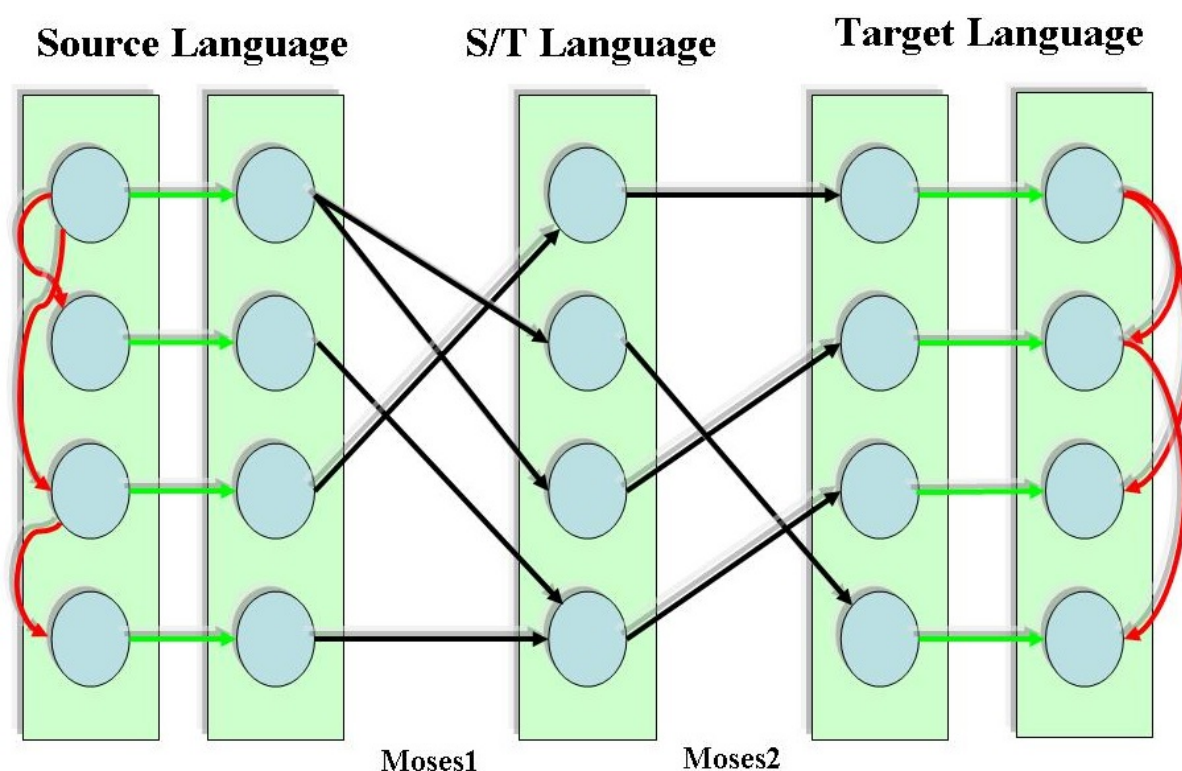

Figure 1: A Hybrid Architecture for Transferring of Linguistic Information from the Source to the Target Language. The linguistic analyses for the source language are projected to a tokinized source text; then Moses models are applied for producing a target language translation. The translation alignment is used for transferring the information to the corresponding tokens in the target language and is used for post processing.

The NLP preprocessing uses the Core NLP package tool for English. The result from the linguistic processing is stored on token level in the source text. The translation steps are performed by a pipeline of two Moses models (Moses1 and Moses2 in the Fig. 1). The first Moses model is phrase based trained on Europarl, Setimes, LibreOffice parallel corpora. The result is used to form intermediate text (Source/Target Language (S/T) text). The construction of S/L text is based on Word Sense Disambiguation similar to approach described in (Simov et al., 2016): it substitutes the words based on the inferred WordNet synsets (if available) and the substitution rules from English to Bulgarian. They take the most frequent lemma in the Bulgarian synset (calculated against a corpus of 70 million words) doctor#1, doc#1, physician#1, MD#2, Dr.#2, medico#2 = doktor#1, lekar#1, d-r#1. Additionally, some words in

the source language are transferred to the target language lemmas on the basis of the results from Moses model one (Moses1). The second-level transfer is achieved through training Moses (Moses2 in the figure) on the factored, partially translated corpus from the first level transfer. It replaces the word-form with a representative lemma from the WordNet synset in the target language: napredak|progress|nn.

The alignment between the source text and the target text created during the translation with Moses system is used for transferring of linguistic information from the source linguistic analysis to the target translation. Here is an example of such an alignment of the sentence "Place them in the midst of a pile of dirty, wet soccer kit." and its translation into Bulgarian[4]:

```
(place them in) = (postavyaneto im v)
(the midst of) = (razgar na)
(a pile of) = (kup)
(dirty) = (izmyrsyavam)
(,) = (,)
(soccer) = (futbolni)
(kit) = (komplekt)
(.) = (.)
```

The transferred information is then used in the postprocessing phase. It should be noted that the alignment between Source Language to S/T Language and from S/T Language to Target is not one-to-one. It maps sequences from the source language to sequences in the target language and the transfer of the linguistic information from the analysis of the source language to the target is not straightforward. Thus, we use heuristic rules. In practice, about 80% of the correspondences are between one-to-one or two-to-two tokens. This facilitates the transfer of the information. But the transferred information is only partial. For the definition of the rules we exploit also language resources and tools for the target language.

In the experiments we use four sets of parallel data (QTLeap corpus: Batch1 to Batch4) and several versions of the translation systems (Baseline, Pilot1, Pilot2 and Pilot3). In this hybrid architecture we use the baseline as Moses1 model and Pilot2 with an extended WordNet as Moses2 one. The current results from for EN-to-BG on Batch4 (answers part) of the QTLeap corpus are as follows: Baseline: 19.91 BLEU (no factors); Pilot2: 20.24 BLEU.

The generation phase relies completely on the Moses system. Then, in the post processing step, a rule-based system is applied, based on lingusitically-enhanced information. It includes various types of rules: morphological, syntactic and semantic. An example of a syntactic rule is the transformation of the English NN compounds into the appropriate syntactic structures in Bulgarian. The direct transfer is rare, since the NN compounds are not so frequent in Bulgarian. They would hold for phrases like "business meeting", for example. An interesting fact in this type is the transposition of N1 and N2 from English to Bulgarian in the N2 N1 variant. It happens when N1 is a named entity. For example, "Rila mountain" is translated as "planina Rila" (mountain Rila). The more frequent types are: adjective noun (AN) and noun with a prepositional phrase (NpN).

3 Deep Semantic Processing

In order to fulfil the demands of MT, deep semantic processing must have at its disposal a considerable amount of *semantic language resources* such as syntax/semantic treebanks (DeepBank, Prague Dependency Treebank, PropBank, Groningen Meaning Bank, etc.), semantic lexicons (WordNet, Ontology-based Lexicons, Valency Lexicons, etc.), and background knowledge (ontologies, linked open data, etc.), *which complement the semantic content of the text in considerable depth and scope*. This definition allows for many approaches to semantic processing to be considered as deep semantic processing.

Deep semantic processing might be and in most cases is still language dependent to a great extent. For instance, the predicates involved in the analyses could be based on the lemmas of the word forms in the

---

[4]Please note that the examples from Bulgarian are presented in their transliterated equivalents.

sentences. The addition of background knowledge provides some language independent elements in the semantic content of the text and we hope that this ensures a better semantic transfer.

In this section we present the main parameters behind the Minimal Recursion Semantics (MRS) as an example of underspecified semantic formalisms. MRS underspecifies scope ambiguities for quantifiers and other scope-bearing elements. The selection of MRS is motivated by several facts: (1) it has already been implemented as part of HPSG grammars for several project languages: English (Copestake and Flickinger, 2000), German (Crysmann, 2007), Spanish (Marimon et al., 2007), Portuguese (Branco and Costa, 2008), (Costa and Branco, 2010) and Bulgarian (Osenova, 2010); (2) it is already used as a basis for semantic transfer in MT systems for several language pairs — (Bond et al., 2005) and (Oepen et al., 2004); (3) it allows the construction of semantic representation over shallow analyses or dependency syntactic structures — (Copestake, 20042006), (Copestake, 2007); (4) there exist corpora annotated with MRS structures, including some parallel ones — (Flickinger et al., 2012b) and (Flickinger et al., 2012a).

At the same time, it should be noted that the existence of precise and robust linguistic grammars for various languages requires time-and-labour-consuming work, in spite of the attempts of the community to provide various start-up kits for better enhancement of new languages. Thus, at the moment the ERG grammar for English is well developed and ready for real applications. Close to it are the grammars for Japanese, Spanish, Norwegian, etc., but in general the current available grammars for many other languages seem to be under-developed and toy-suited. Another issue is that it is challenging to produce full MRSes from dependency parses. Thus, the concept of RMRS is more suitable for real MT applications, in which the production of the logical forms is partial.

3.1 Minimal Recursion Semantics Representation

3.1.1 MRS Definition

MRS is introduced as an underspecified semantic formalism (Copestake et al., 2005). It is used to support semantic analyses in the HPSG English Resource Grammar — (Copestake and Flickinger, 2000), but also in other grammar formalisms like LFG. The main idea is the formalism to rule out spurious analyses resulting from the representation of logical operators and the scope of quantifiers. Spurious analyses of logical form over an utterance could be result from different NLP analyses which produce equivalent logical but syntactically different expressions, like the following two formulas: $\lambda x[\text{fierce}(x) \wedge (\text{black}(x) \wedge \text{cat}(x))]$ and $\lambda x[\text{cat}(x) \wedge (\text{black}(x) \wedge \text{fierce}(x))]$. In MRS such spurious analyses are excluded by the flat representation of the body in the formulas. The determination of the scope of quantifiers in a sentence very often requires information which is not available during the sentence processing. Thus, MRS provides a compact representation which allows further specialization of the quantifiers scope when the necessary information becomes available.

Here we will present only basic definitions from (Copestake et al., 2005). For more details the cited publication should be consulted. An MRS structure is a tuple $< GT, R, C >$, where GT is the top handle, R is a bag of EPs (elementary predicates) and C is a bag of handle constraints, such that there is no handle h that outscopes GT. Each elementary predication contains exactly four components: (1) a handle which is the label of the EP; (2) a relation; (3) a list of zero or more ordinary variable arguments of the relation; and (4) a list of zero or more handles corresponding to scopal arguments of the relation (i.e., holes). Each scopal argument could be assigned to a handle. In this way the elementary predicates with this handle become arguments of other elementary predicates. For example, a quantifier requires such kind of arguments as a body or restriction of the quantifier. Thus handles are used to represent different readings of the same set of elementary predicates via different assignments from handles to the corresponding arguments — see the example below, taken from the cited paper. The handle constraints have the following form: $h_i = h_j$ which states that h_i outscopes h_j. Here is an example of a complex MRS structure for the sentence "Every dog chases some white cat."

$$< h_0, \{h_1 : every(x, h_2, h_3), h_2 : dog(x), h_4 : chase(x, y), h_5 : some(y, h_6, h_7),$$

$$h_6 : white(y), h_6 : cat(y)\}, \{\} >$$

24

The top handle is h_0. The two quantifiers are represented as relations $every(x, y, z)$ and $some(x, y, z)$ where x is the bound variable, y and z are handles determining the restriction and the body of the quantifier. The conjunction of two or more relations is represented by sharing the same handle (h_6 above). The outscope relation is defined as a transitive closure of the immediate outscope relation between two elementary predications — EP immediately outscopes EP' iff one of the scopal arguments of EP is the label of EP'. In this example the set of handle constraints is empty, which means that the representation is underspecified with respect to the scope of both quantifiers. The representation of the MRS structure via bags of elementary predicates and handle constraints allows an easy definition of compositional rules via union of bags and variable substitution. This feature of the MRS provides an easy mechanism for implementation of MRS processors in combination with a syntactic parsing.

3.1.2 RMRS Definition

Robust Minimal Recursion Semantics (RMRS) is introduced as a modification of MRS which captures the semantics resulting from a shallow analysis — see (Copestake, 20042006) and (Copestake, 2007). The main motivations for this development are the facts that currently no single system can do everything: both deep and shallow processing have inherent strengths and weaknesses; on the other hand, the domain-dependent and domain-independent processing must be linked. The ideal level on which this linking can take place is semantics. Therefore, (Copestake, 20042006) and (Copestake, 2007) propose a semantic representation which allows to build a comparable semantic representation for both deep and shallow processing. The justification behind RMRS is to add more underspecification to MRS. This is done by e.g. the separation of arguments from the predicates. Thus each predicate is represented via its name (constructed on the basis of the lemma of the word form in the text) and its main argument which depends on the part of speech - *referential index* for nouns and some pronouns or *event index* in other cases. In this way it is possible that the predicates and their arguments are added to the structure separately from each other. Here we present a formal definition of RMRS as defined in (Jakob et al., 2010). An RMRS structure is a quadruple

$$< hook, EPbag, argumentset, handleconstraints >$$

where a hook consists of three elements $l : a : i$, l is a label, a is an anchor and i is an index. Each elementary predication is additionally marked with an anchor[5] — $l : a : r(i)$, where l is a label, a is an anchor and $r(i)$ is a relation with one argument of appropriate kind — referential index or event index. The argument set contains argument statements of the following kind $a : ARG(x)$, where a is anchor which determines for which relation the argument is defined, ARG is the name of the argument, and x is an index or a hole variable or handle (h) for scopal predicates. The handle constraints are of the form $h =_q l$, where h is a handle, l is a label and $=_q$ is the relation expressing the constraint similarly to MRS. $=_q$ sometimes is written as qeq.

Both representations MRS and RMRS could be transferred to each other under certain conditions. For MRS-to-RMRS it will be necessary to have access to the word forms in the text from which the corresponding predicates were inferred. For RMRS-to-MRS it will be necessary to unify the number of arguments of predicates via some kind of a lexicon. The separation of the predicates from their arguments facilitates the construction of RMRS structures over shallow analyses. Shallow processors usually do not have access to a lexicon. Thus they cannot predict the amount of the arguments that have the corresponding predicate. The forming of the relation names follows such conventions that provide possibilities to construct a correct semantic representation only on the base of information provided by a POS tagger, for example.

3.2 MRS Processing

As it was mentioned above, the RMRS analyses generate over partial and shallow analyses. The idea is to extract as much as possible semantic information from a partial or shallow processed text. In the worst

---

[5]The anchors determine the tokens which generate the corresponding elementary predicates and related arguments. This information facilitates the transfer of information from the source text to target one.

case — from POS tagged text. (Copestake, 2007) demonstrates how RMRS structures can be constructed over the output of a robust statistical parser RASP, which does not have access to subcategorisation information (Briscoe and Carroll, 2002).

The input for the RMRS structures module is based on the following linguistic annotation — the lemma ($Lemma$) for the given wordform; the morphosyntactic tag ($MSTag$) of the wordform, and the dependent relations (Rel) in the dependency tree. In cases of quantifiers we have access to the lexicon used in the Bulgarian HPSG grammar. The algorithm for producing of RMRS from a dependency parse is implemented via two types of rules:

$$< Lemma, MSTag > \rightarrow EP - RMRS$$

The rules of this type produce an RMRS structure representing an elementary predicate.

$$< DRMRS, Rel, HRMRS > HRMRS'$$

The rules of this type unite the RMRS constructed for a dependent node ($DRMRS$) into the current RMRS for a head node ($HRMRS$). The union ($HRMRS'$) is determined by the dependency relation (Rel) between the two nodes.

First, we start with assigning EPs for each lemma in the dependency tree. These EPs are similar to node EPs of (Jakob et al., 2010). Each EP for a given lemma consists of a predicate generated on the basis of the lemma string. Additionally, the morphosyntactic features of the wordform are presented. On the basis of the part-of-speech tag the type of ARG0 is determined — referential index or event index. After this initial step the basic RMRS structure for each lemma in the sentence is compiled. Then these structures are incorporated in each other in bottom-up manner. Here are examples of two RMRS structures constructed in this way. They are in Bulgarian: (1) an RMRS for the verb *'cheta'* (to read):

$< l1 : a1 : e1, \{l1 : a1 :$cheta$\_v\_rel(e1)\}, \{a1 : ARG1(x1)\}, \{\} >$

In this example we also include information for the unexpressed subject (ARG1) which is always incorporated in the verb form. The RMRS structure for a sentence with an explicit subject and an explicit direct object follows. The sentence is *momche mu chete kniga* [Boy him-dative reads book], 'A boy reads a book to him')[6]:

$< l2 : a3 : e1,$
$\{l1 : a1 :$momche$\_n\_rel(x1), l2 : a3 :$chete$\_v\_rel(e1), l3 : a4 :$kniga$\_n\_rel(x2)\},$
$\{a3 : ARG1(x1), a3 : ARG2(x2), a3 : ARG3(x3)\},$
$\{\} >$

The construction of RMRS for sentences, based on shallow processing, suffers from the pipeline processing effect — error accumulation. Errors in earlier processing stages cause suboptimal performance during the next steps.

3.3 MRS for Deep Semantic Transfer in MT

Minimal Recursion Semantics was originally developed for the purposes of Machine Translation. The main idea was that an underspecified semantic representation is appropriate for machine translation because it provides an abstract level to semantic transfer, but at the same time it postpones the difficult decisions. These difficulties are assumed to be less important in the area of machine translation. MRS was applied in the past in two ways to support machine translation: (1) rule-based semantic transfer, and (2) factor-based statistical machine translation. The rules-based semantic transfer uses transfer rules working on the MRS representation of the source language MRS structures and constructing the target language MRS structure. Thus, the transfer rules in this framework are rewriting rules over MRS (Minimal Recursion Semantics) structures. The basic format of the transfer rules is:

$[\mathcal{C} :]\mathcal{I}[!\mathcal{F}] \rightarrow \mathcal{O}$

where $\mathcal{I}$ is the *input* of the rule, $\mathcal{O}$ is the *output*. $\mathcal{C}$ determines the *context* and $\mathcal{F}$ is the *filter* of the rule. $\mathcal{C}$ selects the positive and $\mathcal{F}$ the negative context for the application of a rule. For more details on

---

[6]In this case the information coming from clitics is represented only in the argument set.

the transfer rules, see (Oepen, 2008). This type of rules allows an extremely flexible transfer of factual and linguistic knowledge between the source and the target languages. The rules have access to the MRS structure for the source language, but can also access the (partially) constructed target language MRS structure. Thus elements in each rule could include parts from both MRS structures. This approach requires a good deep grammar for the source language which produces complete MRS structures, then a complete set of rules for transferring of the source language MRSes to the target language MRSes and a generation grammar for the target language. As already discussed, there are no many languages equipped with such grammars and rules.

The factor-based translation model is built on top of the factored SMT model proposed by (Koehn and Hoang, 2007), as an extension of the traditional phrase-based SMT framework. Instead of using only the word form of the text, it allows the system to take a vector of factors to represent each token, both for the source and target languages. The vector of factors can be used for different levels of linguistic annotations, like lemma, part-of-speech, or other linguistic features, if they can be (somehow) represented as annotations to each token.

In our set-up we have used factored-based transfer of elementary predicates from MRSes in our earlier systems. The results were only slightly better or slightly worse than the baseline system. Thus we decided to use the hybrid system described above in order to transfer deep information from the analysis of the source language to the automatic translation in the target language where this information, together with the language resources for the target language is used for a post processing over the automatic translation.

We are still implementing rules for our hybrid approach, thus, the results we report here, are only preliminary. We have implemented transfer of the grammatical features from the source to the target language, processing of complex NE and NN compounds. The current result is 21.78 BLEU.

4 Conclusions

In this paper a hybrid architecture for deep MT was presented in the language direction from English to Bulgarian. The implementation provided some relaxation on the translation model, suggesting a two-level transfer model. The former being semantically coarse, on token level, but already using a WSD constraint; the latter being more fine-grained, on lemma level. Instead of full MRS approach, we used the RMRS variant, encoded in the factored MT model. In the post processing step some frequent errors have been corrected on morhological, syntactic and semantic levels. However, this is ongoing work. The experiments showed that this approach adds to the BLEU score results and thus is promising for further elaboration.

The actual translation approach could be different from the one presented here. Any translation system that supports alignment could be used. For example, as one of the reviewers suggested, neural network machine translation with attention could be also used. This is one direction for future work.

Acknowledgements

References

Francis Bond, Stephan Oepen, Melanie Siegel, Ann Copestake, and Dan Flickinger. 2005. Open source machine translation with DELPH-IN. In *Proceedings of the Open-Source Machine Translation Workshop at the 10th Machine Translation Summit*, pages 15 – 22.

Johan Bos. 2013. The groningen meaning bank. In *Invited Talk at Joint Symposium on Semantic Processing: Textual Inference and Structures in Corpora*.

Antonia Branco and Francisco Costa. 2008. LXGram in the Shared Task "Comparing Semantic Representations" of STEP 2008. In Johan Bos and Rodolfo Delmonte, editors, *Semantics in Text Processing. STEP 2008 Conference Proceedings*, volume 1 of *Research in Computational Semantics*, pages 299–314. College Publications.

Ted Briscoe and John Carroll. 2002. Robust accurate statistical annotation of general text. In *Proceedings of LREC-2002*, May. ACL Anthology Identifier: L02-1250.

Ann Copestake and Dan Flickinger. 2000. An open source grammar development environment and broad-coverage english grammar using hpsg. In *Proceedings of LREC-2000*.

Ann Copestake, Dan Flickinger, Carl Pollard, and Ivan Sag. 2005. Minimal recursion semantics: An introduction. *Research on Language & Computation*, 3(4):281–332.

Ann Copestake. 2004/2006. Robust minimal recursion semantics. unpublished draft.

Ann Copestake. 2007. Applying robust semantics. In *Proceedings of the 10th Conference of the Pacific Assocation for Computational Linguistics (PACLING)*, pages 1–12.

Francisco Costa and António Branco. 2010. Lxgram: A deep linguistic processing grammar for Portuguese. In *Lecture Notes in Artificial Intelligence*, volume 6001, pages 86–89. Springer, Berlin, May.

Berthold Crysmann. 2007. Local ambiguity packing and discontinuity in german. In *ACL 2007 Workshop on Deep Linguistic Processing*, pages 144–151.

Daniel Flickinger, Valia Kordoni, Yi Zhang, Antnio Branco, Kiril Simov, Petya Osenova, Catarina Carvalheiro, Francisco Costa, and Srgio Castro. 2012a. Pardeepbank: Multiple parallel deep treebanking. In *Proceedings of TLT-11*, pages 97–108. Edies Colibri, Lisbon.

Daniel Flickinger, Yi Zhang, and Valia Kordoni. 2012b. Deepbank: A dynamically annotated treebank of the wall street journal. In *Proceedings of TLT-11*, pages 85–96. Edies Colibri, Lisbon.

Jan Hajič. 2011. Machine translation research in META-NET. presentation at META-NET meeting.

Max Jakob, Markéta Lopatková, and Valia Kordoni. 2010. Mapping between dependency structures and compositional semantic representations. In *Proceedings of LREC 2010*, pages 2491–2497.

Paul Kingsbury and Martha Palmer. 2002. From treebank to propbank. In *Proceedings of LREC-2002*, pages 1989–1993.

Philipp Koehn and Hieu Hoang. 2007. Factored translation models. In *Proceedings of EMNLP*.

Montserrat Marimon, Núria Bel, Sergio Espeja, and Natalia Seghezzi. 2007. The spanish resource grammar: Pre-processing strategy and lexical acquisition. In *ACL 2007 Workshop on Deep Linguistic Processing*, pages 105–111.

Stephan Oepen, Helge Dyvik, Jan Tore Lnning, Erik Velldal, Dorothee Beermann, John Carroll, Dan Flickinger, Lars Hellan, Janne Bondi Johannessen, Paul Meurer, Torbjrn Nordgrd, and Victoria Rosn. 2004. Som kappete med trollet? Towards MRS-based Norwegian – English Machine Translation. In *Proceedings of the 10th International Conference on Theoretical and Methodological Issues in Machine Translation*.

Stephan Oepen. 2008. The Transfer Formalism. General Purpose MRS Rewriting. Technical Report LOGON Project. Technical report, University of Oslo.

Petya Osenova. 2010. *The Bulgarian Resource Grammar*. VDM.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The proposition bank: An annotated corpus of semantic roles. *Comput. Linguist.*, 31(1):71–106, March.

Kiril Simov, Petya Osenova, and Alexander Popov. 2016. Towards semantic-based hybrid machine translation between bulgarian and english. In *Proceedings of the 2nd Workshop on Semantics-Driven Machine Translation (SedMT 2016)*, pages 22–26, San Diego, California, June. Association for Computational Linguistics.