

Towards Domain-Independent Deep Linguistic Processing: Ensuring Portability and Re-Usability of Lexicalised Grammars

Kostadin Cholakov[†], Valia Kordoni^{†‡}, Yi Zhang^{†‡}

[†] Department of Computational Linguistics, Saarland University, Germany

[‡] LT-Lab, DFKI GmbH, Germany

{kostadin,kordoni,yzhang}@coli.uni-sb.de

Abstract

In this paper we illustrate and underline the importance of making detailed linguistic information a central part of the process of automatic acquisition of large-scale lexicons as a means for enhancing robustness and at the same time ensuring maintainability and re-usability of *deep* lexicalised grammars. Using the error mining techniques proposed in (van Noord, 2004) we show very convincingly that the main hindrance to portability of *deep* lexicalised grammars to domains other than the ones originally developed in, as well as to robustness of systems using such grammars is low lexical coverage. To this effect, we develop linguistically-driven methods that use detailed morphosyntactic information to automatically enhance the performance of *deep* lexicalised grammars maintaining at the same time their usually already achieved high linguistic quality.

1 Introduction

We focus on enhancing robustness and ensuring maintainability and re-usability for a large-scale *deep* grammar of German (GG; (Crysmann, 2003)), developed in the framework of Head-driven Phrase Structure Grammar (HPSG). Specifically, we show that the incorporation of detailed linguistic information into the process of automatic extension of the lexicon of such a language resource enhances its performance and provides linguistically sound and more informative predictions which bring a bigger benefit for the grammar when employed in practical real-life applications.

In recent years, various techniques and resources have been developed in order to improve robustness of deep grammars for real-life applications in various domains. Nevertheless, *low coverage* of such grammars remains the main hindrance to their employment in open domain natural language processing. (Baldwin et al., 2004), as well as (van Noord, 2004) and (Zhang and Kordoni, 2006) have clearly shown that the majority of parsing failures with large-scale deep grammars are caused by missing or wrong entries in the lexicons accompanying grammars like the aforementioned ones. Based on these findings, it has become clear that it is crucial to explore and develop efficient methods for *automated (Deep) Lexical Acquisition* (henceforward (D)LA), the process of automatically recovering missing entries in the lexicons of deep grammars.

Recently, various high-quality DLA approaches have been proposed. (Baldwin, 2005), as well as (Zhang and Kordoni, 2006), (van de Cruys, 2006) and (Nicholson et al., 2008) describe efficient methods towards the task of lexicon acquisition for large-scale deep grammars for English, Dutch and German. They treat DLA as a classification task and make use of various robust and efficient machine learning techniques to perform the acquisition process.

However, it is our claim that to achieve better and more practically useful results, apart from good learning algorithms, we also need to *incorporate* into the learning process fine-grained linguistic information which deep grammars inherently include and provide for. As we clearly show in the following, it is **not** sufficient to only develop and use good and complicated classification algorithms. We must look at the detailed linguistic information that is already included and provided for by the grammar itself and try to capture and make as much use of it as possible, for this is the information we aim at learning when performing DLA.

© 2008. Licensed under the *Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported* license (<http://creativecommons.org/licenses/by-nc-sa/3.0/>). Some rights reserved.

In this way, the learning process is facilitated and at the same time it is as much as possible ensured that its outcome be linguistically more informative and, thus, practically more useful.

We use the GG deep grammar for the work we present in this paper because German is a language with rich morphology and free word order, which exhibits a range of interesting linguistic phenomena, a fair number of which are already analysed in the GG. Thus, the grammar is a valuable linguistic resource since it provides linguistically sound and detailed analyses of these phenomena. Apart from the interesting syntactic structures, though, the lexical entries in the lexicon of the aforementioned grammar also exhibit a rich and complicated structure and contain various important linguistic constraints. Based on our claim above, in this paper we show how the information these constraints provide can be captured and used in *linguistically-motivated* DLA methods which we propose here. We then apply our approach on real-life data and observe the impact it has on the the grammar coverage and its practical application. In this way we try to prove our assumption that the linguistic information we incorporate into our DLA methods is vital for the good performance of the acquisition process and for the maintainability and re-usability of the grammar, as well for its successful practical application.

The remainder of the paper is organised as follows. In Section 2 we show that low (lexical) coverage is a serious issue for the GG when employed for open domain natural language processing. Section 3 presents the types in the lexical architecture of the GG that are considered to be relevant for the purposes of our experiments. Section 4 describes the extensive linguistic analysis we perform in order to deal with the linguistic information these types provide and presents the target type inventory for our DLA methods. Section 5 reports on statistical approaches towards automatic DLA and shows the importance of a good and linguistically-motivated feature selection. Section 6 illustrates the practical usage of the proposed DLA methods and their impact on grammar coverage. Section 7 concludes the paper.

2 Coverage Test with the GG

We start off adopting the automated error mining method described in (van Noord, 2004) for identification of the major type of errors in the GG.

As an HPSG grammar, the GG is based on typed feature structures. The GG types are strictly defined within a type hierarchy. The GG also contains constructional and lexical rules and a lexicon with its entries belonging to lexical types which are themselves defined again within the type hierarchy. The grammar originates from (Müller and Kasper, 2000), but continued to improve after the end of the Verbmobil project (Wahlster, 2000) and it currently consists of 5K types, 115 rules and the lexicon contains approximately 35K entries. These entries belong to 386 distinct lexical types.

In the experiments we report here two corpora of different kind and size have been used. The first one has been extracted from the Frankfurter Rundschau newspaper and contains about 614K sentences that have between 5 and 20 tokens. The second corpus is a subset of the German part of the Wacky project (Kilgariff and Grefenstette, 2003). The Wacky project aims at the creation of large corpora for different languages, including German, from various web sources, such as online newspapers and magazines, legal texts, internet fora, university and science web sites, etc. The German part, named deWaC (Web as Corpus), contains about 93M sentences and 1.65 billion tokens. The subset used in our experiments is extracted by randomly selecting 2.57M sentences that have between 4 and 30 tokens. These corpora have been chosen because it is interesting to observe the grammar performance on a relatively balanced newspaper corpus that does not include so many long sentences and sophisticated linguistic constructions and to compare it with the performance of the grammar on a random open domain text corpus.

The sentences are fed into the PET HPSG parser (Callmeier, 2000) with the GG loaded. The parser has been configured with a maximum edge number limit of 100K and it is running in the *best-only* mode so that it does not exhaustively find all possible parses. The result of each sentence is marked as one of the following four cases:

- *P* means at least one parse is found for the sentence;
- *L* means the parser halted after the morphological analysis and was not able to construct any lexical item for the input token;
- *N* means that the parser exhausted the searching and was not able to parse the sentence;

- *E* means the parser reached the maximum edge number limit and was still not able to find a parse.

Table 1 shows the results of the experiments with the two corpora. From these results it can

Result	FR		deWaC	
	#Sentences	%	#Sentences	%
<i>P</i>	62,768	10.22%	109,498	4.3%
<i>L</i>	464,112	75.55%	2,328,490	90.5%
<i>N</i>	87,415	14.23%	134,917	5.2%
<i>E</i>	3	–	14	–
Total:	614,298	100%	2,572,919	100%

Table 1: Parsing results with the GG and the test corpora

be seen that the GG has full lexical span for only a small portion of the sentences— about 25% and 10% for the Frankfurter Rundschau and the deWaC corpora, respectively. The output of the error mining confirms our assumption that missing lexical entries are the main problem when it comes to robust performance of the GG and illustrates the need for efficient DLA methods.

3 Atomic Lexical Types

Before describing the proposed DLA algorithm, we should define what exactly is being learnt. Most of the so called *deep* grammars are strongly lexicalised. As mentioned in the previous section, the GG employs a type inheritance system and its lexicon has a flat structure with each lexical entry mapped onto one type in the inheritance hierarchy. Normally, the types assigned to the lexical entries are maximal on the type hierarchy, i.e., they do not have any subtypes. They provide the most specific information available for this branch of the hierarchy. These maximal types which the lexical entries are mapped onto are called *atomic lexical types*. Thus, in our experiment setup, we can define the lexicon of the grammar as being a one-to-one mapping from word stems to atomic lexical types. It is this mapping which must be automatically learnt (guessed) by the different DLA methods.

We are interested in learning open-class words, i.e., nouns, adjectives, verbs and adverbs. We assume that the close-class words are already in the lexicon or the grammar can handle them through various lexical rules and they are not crucial for the grammar performance in real life applications. Thus, for the purposes of our experiments, we consider only the open-class lexical types. Moreover,

we propose an *inventory of open-class lexical types with sufficient type and token frequency*. The **type frequency** of a given lexical type is defined as the number of lexical entries in the lexicon of the grammar that belong to this type and the **token frequency** is the number of words in some corpus that belong to this type.

We use sentences from the Verbmobil corpus which have been treebanked with the GG in order to determine the token frequency and to map the lexemes to their correct entries in the lexicon for the purposes of the experiment. This set contains 11K sentences and about 73K tokens; this gives an average of 6.8 words per sentence. The sentences are taken from spoken dialogues. Hence, they are not long and most of them do not exhibit interesting linguistic properties which is a clear drawback but currently there is no other annotated data compatible with the GG.

We used a type frequency threshold of 10 entries in the lexicon and a token frequency threshold of 3 occurrences in the treebanked sentences to form a list of relevant open-class lexical types. The resulting list contains 38 atomic lexical types with a total of 32,687 lexical entries.

4 Incorporation of Linguistic Features

However, in the case of the GG this type inventory is not a sufficient solution. As already mentioned, in the lexicon of the grammar much of the relevant linguistic information is encoded not in the type definition itself but in the form of constraints in the feature structures of the various types. Moreover, given that German has a rich morphology, a given attribute may have many different values among lexical entries of the same type and it is crucial for the DLA process to capture all the different combinations. That is why we expand the identified 38 atomic lexical type definitions by including the values of various features into them.

By doing this, we are trying to facilitate the DLA process because, in that way, it can ‘learn’ to differentiate not only the various lexical types but also significant morphosyntactic differences among entries that belong to the same lexical type. That gives the DLA methods access to much more linguistic information and they are able to apply more linguistically fine-tuned classification criteria when deciding which lexical type the unknown word must be assigned to. Furthermore, we ensure that the learning process deliver linguistically

Feature	Values	Meaning
SUBJOPT (subject options)	+	in some cases the article for the noun can be omitted
	-	the noun always goes with an article
	+	raising verb
	-	non-raising verb
KEYAGR (key agreement)	-	case-number-gender information for nouns
	c-s-n	underspecified-singular-neutral
	c-p-g	underspecified-plural-underspecified

(O)COMPAGR ((oblique) complement agreement)	a-n-g, d-n-g, etc.	case-number-gender information for (oblique) verb complements
	-	case-number-gender of the modified noun (for adjectives)
	-	for (oblique) verb complements
(O)COMPTOPT ((oblique) complement options)	-	verbs can take a different number of complements
	+	the respective (oblique) complement is present
	-	the respective (oblique) complement is absent
KEYFORM	-	the auxiliary verb used for the formation of perfect tense
	haben	the auxiliary verb is 'haben'
	sein	the auxiliary verb is 'sein'

Table 2: Relevant features used for type expansion

plausible, precise and more practically useful results. The more the captured and used linguistic information is, the better and more useful the DLA results will be.

However, we have to avoid creating data sparse problems. We do so by making the assumption that not every feature could really contribute to the classification process and by filtering out these features that we consider *irrelevant* for the enhancement of the DLA task. Naturally, the question which features are to be considered relevant arises. After performing an extensive linguistic analysis, we have decided to take the features shown in Table 2 into account.

We have thoroughly analysed each of these features and selected them on the basis of their linguistic meaning and their significance and contribution to the DLA process. The SUBJOPT feature can be used to differentiate among nouns that have a similar morphosyntactic behaviour but differ only in the usage of articles; 4 out of the considered 9 noun atomic lexical types do not define this feature. Furthermore, using this feature, we can also refine our classification within a single atomic lexical type. For example, the entry '*adresse-n*' (address) of the type 'count-noun-le'¹ has '-' for the SUBJOPT value, whereas the value for the entry '*anbindung-n*' (connection) of the same type is '+':

- (1) a. Das Hotel hat gute
det.NEUT.NOM hotel have.3PER.SG good
Anbindung an die öffentlichen
connection to det.PL.ACC public

¹count noun lexeme; all lexical entries in the lexicon end with *le* which stands for lexeme.

Verkehrsmittel.
transportation means
'The hotel has a good connection to public transportation.'

- b. **Die** *Anbindung* an Rom mit
det.FEM.NOM connection to Rome with
dem Zug ist gut.
det.MASC.DAT train be.3PER.SG good
'The train connection to Rome is good.'

The distinction between raising and non-raising verbs that this feature expresses is also an important contribution to the classification process.

The case-number-gender data the KEYAGR and (O)COMPAGR features provide allows for a better usage of morphosyntactic information for the purposes of DLA. Based on this data, the classification method is able to capture words with similar morphosyntactic behaviour and give various indications for their syntactic nature; for instance, if the word is a subject, direct or indirect object. This is especially relevant and useful for languages with rich morphology and relatively free word order such as German. The same is also valid for the (O)COMPTOPT and KEYFORM features– they allow the DLA method to successfully learn and classify verbs with similar syntactic properties.

The values of the features are just attached to the old type name to form a new type definition. In this way, we 'promote' them and these features are now part of the type hierarchy of the grammar which makes them accessible for the DLA process since this operates on the type level. For example, the original type of the entry for the noun 'abenteuer' (adventure):

abenteuer-n := count-noun-le &
[[--SUBJOPT -,

```

KEYAGR c-n-n,
KEYREL "_abenteuer_n_rel",
KEYSORT situation,
MCLASS nclass-2_-u_-e ] ].

```

will become *abenteuer-n := count-noun-le_-c-n-n* when we incorporate the values of the features SUBJOPT and KEYAGR into the original type definition. The new expanded type inventory is shown in Table 3.

	Original lexicon	Expanded lexicon
Number of lexical types	386	485
Atomic lexical types	38	137
-nouns	9	72
-verbs	19	53
-adjectives	3	5
-adverbs	7	7

Table 3: Expanded atomic lexical types

The features we have ignored do not contribute to the learning process and are likely to create sparse data problems. The (O)COMPFORM ((oblique) complement form) features which denote dependent to verbs prepositions are not considered to be relevant. An example of OCOMPFORM is the lexical entry ‘begründen_mit-v’ (justify with) where the feature has the preposition ‘mit’ (with) as its value. Though for German prepositions can be considered as case markers, the DLA has already a reliable access to case information through the (O)COMPAGR features. Moreover, a given dependent preposition is distributed across many types and it does not indicate clearly which type the respective verb belongs to.

The same is valid for the feature VCOPMFORM (verb complement form) that denotes the separable particle (if present) of the verb in question. An example of this feature is the lexical entry ‘abdecken-v’ (to cover) where VCOPMFORM has the separable particle ‘ab’ as its value. However, treating such discontinuous verb-particle combinations as a lexical unit could help for the acquisition of subcategorizational frames. For example, *anhören* (to listen to someone/something) takes an accusative NP as argument, *zuhören* (to listen to) takes a dative NP and *aufhören* (to stop, to terminate) takes an infinitival complement. Thus, ignoring VCOPMFORM could be a hindrance for the acquisition of some verb types².

We have also tried to incorporate some sort of semantic information into the expanded atomic

²We thank the anonymous reviewer who pointed this out for us.

lexical type definitions by also attaching the KEYSORT semantic feature to them. KEYSORT defines a certain situation semantics category (‘anything’, ‘action_sit’, ‘mental_sit’) which the lexical entry belongs to. However, this has caused again a sparse data problem because the semantic classification is too specific and, thus, the number of possible classes is too large. Moreover, semantic classification is done based on completely different criteria and it cannot be directly linked to the morphosyntactic features. That is why we have finally excluded this feature, as well.

Armed with this elaborate target type inventory, we now proceed with the DLA experiments for the GG.

5 DLA Experiments with the GG

For our DLA experiments, we adopted the Maximum Entropy based model described in (Zhang and Kordoni, 2006), which has been applied to the ERG (Copestake and Flickinger, 2000), a wide-coverage HPSG grammar for English. For the proposed prediction model, the probability of a lexical type t given an unknown word and its context c is:

$$(2) \quad p(t|c) = \frac{\exp(\sum_i \Theta_i f_i(t,c))}{\sum_{t' \in T} \exp(\sum_i \Theta_i f_i(t',c))}$$

where $f_i(t, c)$ may encode arbitrary characteristics of the context and Θ_i is a weighting factor estimated on a training corpus. Our experiments have been performed with the feature set shown in Table 4.

Features
the prefix of the unknown word (length is less or equal 4)
the suffix of the unknown word (length is less or equal 4)
the 2 words before and after the unknown word
the 2 types before and after the unknown word

Table 4: Features for the DLA experiment

We have also experimented with prefix and suffix lengths up to 3. To evaluate the contribution of various features and the overall precision of the ME-based unknown word prediction model, we have done a 10-fold cross validation on the VerbMobil treebanked data. For each fold, words that do not occur in the training partition are assumed to be unknown and are temporarily removed from the lexicon.

For comparison, we have also built a baseline model that always assigns a majority type to each

unknown word according to its POS tag. Specifically, we tag the input sentence with a small POS tagset. It is then mapped to a most popular lexical type for that POS. Table 5 shows the relevant mappings.

POS	Majority lexical type
noun	count-noun-le-_-c-n-f
verb	trans-nerg-str-verb-le_haben-auxf
adj	adj-non-prd-le
adv	intersect-adv-le

Table 5: POS tags to lexical types mapping

Again for comparison, we have built another simple baseline model using the *TnT* POS tagger (Brants, 2000). *TnT* is a general-purpose HMM-based trigram tagger. We have trained the tagging models with all the lexical types as the tagset. The tagger tags the whole sentence but only the output tags for the unknown words are taken to generate lexical entries and to be considered for the evaluation. The precisions of the different prediction models are given in Table 6.

The baseline achieves a precision of about 38% and the POS tagger outperforms it by nearly 10%. These results can be explained by the nature of the Verbmobil data. The vast majority of the adjectives and the adverbs in the sentences belong to the majority types shown in Table 5 and, thus, the baseline model assigns the correct lexical types to almost every adjective and adverb, which brings up the overall precision. The short sentence length facilitates the tagger extremely, for *TnT*, as an HMM-based tagger, makes predictions based on the whole sentence. The longer the sentences are, the more challenging the tagging task for *TnT* is. The results of these models clearly show that the task of unknown word type prediction for deep grammars is non-trivial.

Our ME-based models give the best results in terms of precision. However, verbs and adverbs remain extremely difficult for classification. The simple morphological features we use in the ME model are not good enough for making good predictions for verbs. Morphology cannot capture such purely syntactic features as subcategorization frames, for example.

While the errors for verbs are pretty random, there is one major type of wrong predictions for adverbs. Most of them are correctly predicted as such but they receive the majority type for adverbs, namely ‘*intersect-adv-le*’. Since most of the adverbs in the Verbmobil data we are using belong

to the majority adverb type, the predictor is biased towards assigning it to the unknown words which have been identified as adverbs.

The results in the top half of the Table 6 show that morphological features are already very good for predicting adjectives. In contrast with adverbs, adjectives occur in pretty limited number of contexts. Moreover, when dealing with morphologically rich languages such as German, adjectives are typically marked by specific affixes corresponding to a specific case-number-gender combination. Since we have incorporated this kind of linguistic information into our target lexical type definitions, this significantly helps the prediction process based on morphological features.

Surprisingly, nouns seem to be hard to learn. Apparently, the vast majority of the wrong predictions have been made for nouns that belong to the expanded variants of the lexical type ‘*count-noun-le*’ which is also the most common non-expanded lexical type for nouns in the original lexicon. Many nouns have been assigned the right lexical type except for the gender:

- (3) *Betrieb* (business, company, enterprise)
 prediction: *count-noun-le-_-c-n-n*
 correct type: *count-noun-le-_-c-n-m*

According to the strict *exact-match* evaluate measure we use, such cases are considered to be errors because the predicted lexical type does not match the type of the lexical entry in the lexicon.

The low numbers for verbs and adverbs show clearly that we also need to incorporate some sort of syntactic information into the prediction model. We adopt the method described in (Zhang and Kordoni, 2006) where the disambiguation model of the parser is used for this purpose. We also believe that the kind of detailed morphosyntactic information which the learning process now has access to would facilitate the disambiguation model because the input to the model is linguistically more fine-grained. In another DLA experiment we let PET use the top 3 predictions provided by the lexical type predictor in order to generate sentence analyses. Then we use the disambiguation model, trained on the Verbmobil data, to choose the best one of these analyses and the corresponding lexical entry is taken to be the final result of the prediction process.

As shown in the last line of Table 6, we achieve an increase of 19% which means that in many cases the correct lexical type has been ranked sec-

Model	Precision	Nouns	Adjectives	Verbs	Adverbs
Baseline	37.89%	27.03%	62.69%	33.57%	67.14%
TnT	47.53%	53.76%	74.52%	26.94%	32.68%
ME(affix length=3)	51.2%	48.25%	75.41%	44.06%	44.13%
ME(affix length=4)	54.63%	53.55%	76.79%	47.10%	43.55%
ME + disamb.	73.54%	75%	88.24%	65.98%	65.90%

Table 6: Precision of unknown word type predictors

ond or third by the predictor. This proves that the expanded lexical types improve also the performance of the disambiguation model and allow for its successful application for the purposes of DLA. It also shows, once again, the importance of the morphology in the case of the GG and proves the rightness of our decision to expand the type definitions with detailed linguistic information.³

6 Practical Application

Since our main claim in this paper is that for good and *practically useful* DLA, which at the same time may facilitate robustness and ensure maintainability and re-usability of *deep* lexicalised grammars, we do not only need good machine learning algorithms but also classification and feature selection that are based on an extensive linguistic analysis, we apply our DLA methods to real test data. We believe that due to our expanded lexical type definitions, we provide much more linguistically accurate predictions. With this type of predictions, we anticipate a bigger improvement of the grammar coverage and accuracy for the prediction process delivers much more *linguistically relevant* information which facilitates parsing with the GG.

We have conducted experiments with PET and the two corpora we have used for the error mining to determine whether we can improve coverage by using our DLA method to predict the types of unknown words online. We have trained the predictor on the whole set of treebanked sentences and extracted a subset of 50K sentences from each corpus. Since lexical types are not available for these sentences, we have used POS tags instead as features for our prediction model. Coverage is measured as the number of sentences that received at least one parse and accuracy is measured as the number of sentences that received a *correct* analysis. The results are shown in Table 7.

The coverage for FR improves with more than 12% and the accuracy number remains almost the

³Another reason for this high result is the short average length of the treebanked sentences which facilitates the disambiguation model of the parser.

Parsed Corpus	Coverage	Accuracy
FR with the vanilla version GG	8.89%	85%
FR with the GG + DLA	21.08%	83%
deWaC with the vanilla version GG	7.46%	–
deWaC with the GG + DLA	16.95%	–

Table 7: Coverage results

same. Thus, with our linguistically-oriented DLA method, we have managed to increase parsing coverage and at the same time to preserve the high accuracy of the grammar. It is also interesting to note the increase in coverage for the deWaC corpus. It is about 10%, and given the fact that deWaC is an open and unbalanced corpus, this is a clear improvement. However, we do not measure accuracy on the deWaC corpus because many sentences are not well formed and the corpus itself contains much ‘noise’. Still, these results show that the incorporation of detailed linguistic information in the prediction process contributed to the parser performance and the robustness of the grammar without harming the quality of the delivered analyses.

7 Conclusion

In this paper, we have tackled from a more linguistically-oriented point of view the lexicon acquisition problem for a large-scale deep grammar for German, developed in HPSG. We have shown clearly that missing lexical entries are the main cause for parsing failures and, thus, illustrated the importance of increasing the lexical coverage of the grammar. The target type inventory for the learning process has been developed in a linguistically motivated way in an attempt to capture significant morphosyntactic information and, thus, achieve a better performance and more practically useful results.

With the proposed DLA approach and our elaborate target type inventory we have achieved nearly 75% precision and this way we have illustrated the importance of fine-grained linguistic information for the lexical prediction process. In the end, we have shown that with our linguistically motivated DLA methods, the parsing coverage of the afore-

mentioned deep grammar improves significantly while its linguistic quality remains intact.

The conclusion, therefore, is that it is vital to be able to capture linguistic information and successfully incorporate it in DLA processes, for it facilitates deep grammars and makes processing with them much more robust for applications. At the same time, the almost self-evident portability to new domains and the re-usability of the grammar for open domain natural language processing is significantly enhanced.

The DLA method we propose can be used as an external module that can help the grammar be ported and operate on different domains. Thus, specifically in the case of HPSG, DLA can also be seen as a way for achieving more modularity in the grammar. Moreover, in a future research, the proposed kind of DLA might also be used in order to facilitate the division and transition from a *core deep grammar* with a *core lexicon* towards *subgrammars* with *domain specific lexicons/lexical constraints* in a linguistically motivated way. The use of both these divisions naturally leads to a highly modular structure of the grammar and the system using the grammar, which at the same time helps in controlling its complexity.

Our linguistically motivated approach provides fine-grained results that can be used in a number of different ways. It is a valuable linguistic tool and it is up to the grammar developer to choose how to use the many opportunities it provides.

References

- Baldwin, Timothy, Emily M. Bender, Dan Flickinger, Ara Kim, and Stephan Oepen. 2004. Road-testing the English Resource Grammar over the British National Corpus. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004)*, Lisbon, Portugal.
- Baldwin, Timothy. 2005. Bootstrapping deep lexical resources: Resources for courses. In *Proceedings of the ACL-SIGLEX 2005 Workshop on Deep Lexical Acquisition*, pages 67–76, Ann Arbor, USA.
- Brants, Thorsten. 2000. TnT- a statistical part-of-speech tagger. In *Proceedings of the Sixth Conference on Applied Natural Language Processing ANLP-2000*, Seattle, WA, USA.
- Callmeier, Ulrich. 2000. PET- a platform for experimentation with efficient HPSG processing techniques. In *Journal of Natural Language Engineering*, volume 6(1), pages 99–108.
- Copetake, Ann and Dan Flickinger. 2000. An open-source grammar development environment and broad-coverage English grammar using HPSG. In *Proceedings of the Second conference on Language Resources and Evaluation (LREC 2000)*, Athens, Greece.
- Crysmann, Berthold. 2003. On the efficient implementation of German verb placement in HPSG. In *Proceedings of RANLP 2003*, pages 112–116, Borovets, Bulgaria.
- Kilgarriff, Adam and G Grefenstette. 2003. Introduction to the special issue on the web as corpus. *Computational Linguistics*, 29:333–347.
- Müller, Stephan and Walter Kasper. 2000. HPSG analysis of German. In Wahlster, Wolfgang, editor, *Verbmobil: Foundations of Speech-to-Speech Translation*, pages 238–253. Springer-Verlag.
- Nicholson, Jeremy, Valia Kordoni, Yi Zhang, Timothy Baldwin, and Rebecca Dridan. 2008. Evaluating and extending the coverage of HPSG grammars. In *In proceedings of LREC*, Marrakesh, Morocco.
- van de Cruys, Tim. 2006. Automatically extending the lexicon for parsing. In Huitink, Janneke and Sophia Katrenko, editors, *Proceedings of the Student Session of the European Summer School in Logic, Language and Information (ESSLLI)*, pages 180–191, Malaga, Spain.
- van Noord, Gertjan. 2004. Error mining for wide coverage grammar engineering. In *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL'04), Main Volume*, pages 446–453, Barcelona, Spain.
- Wahlster, Wolfgang, editor. 2000. *Verbmobil: Foundations of Speech-to-Speech Translation*. Artificial Intelligence. Springer.
- Zhang, Yi and Valia Kordoni. 2006. Automated deep lexical acquisition for robust open text processing. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006)*, Genoa, Italy.