The CoreGram Project: A Brief Overview and Motivation

Stefan Müller

Freie Universität Berlin

Keywords: Universal Grammar, Head-Driven Phrase Structure Grammar, Multilingual Grammar Engineering, TRALE

This paper gives a brief overview of the CoreGram project. For a more detailed description of the project, further motivation and comparison with similar enterprises see Müller, 2015a.

1 Overview and Motivation

The goal of the CoreGram project is the development of large scale computer processable grammar fragments of several languages that share a common core. The theoretical framework is Head-Driven Phrase Structure Grammar (HPSG, Pollard and Sag, 1994; Müller, 2007b). Currently we work on the following languages:

- German (Müller, 2007b, 2009b, 2012; Müller and Ørsnes, 2011a; Müller, 2014a)
- Danish (Ørsnes, 2009; Müller, 2009b, 2012; Müller and Ørsnes, 2011a, 2015)
- Persian (Müller, 2010; Müller and Ghayoomi, 2010; Müller, Samvelian and Bonami, In Preparation)
- Maltese (Müller, 2009a)
- Mandarin Chinese (Lipenkova, 2009; Müller and Lipenkova, 2009)
- Yiddish (Müller and Ørsnes, 2011b)
- English (Müller, 2009b, 2012)
- Spanish
- French

For the implementation we use the TRALE system (Meurers, Penn and Richter, 2002; Penn, 2004), which allows for a rather direct encoding of HPSG analyses (Melnik, 2007; Müller, 2015a). The grammars of German, Danish, Persian, Maltese, and Mandarin Chinese are of non-trivial size and can be downloaded at http://hpsg.fu-berlin.de/Projects/CoreGram.html. They are also part of the next version of the Grammix CD-ROM (Müller, 2007a). The grammars of Yiddish and English are toy grammars that are used to verify cross-linguistic analyses of special phenomena and the work on Spanish and French is part of work in the Sonderforschungsbereich 632 which just started. See Bildhauer, 2008 for an implemented grammar of Spanish that will be converted into the format of the grammars mentioned above.

I believe that working out large scale computer-implemented grammars is the best way to verify the consistency of linguistic theories (Müller, 1999, Chapter 22; Bender, 2008). Much linguistic work is published in journal articles but the underlying assumptions of articles may be different so that it is difficult to imagine a coherent view that incorporates all insights. Even for articles by the same author it is not guaranteed that basic assumptions are shared between articles since it can take several years for individual papers until they are published. Hence, I believe that books are the right format for describing linguistic theories and ideally such theories are backed up by computer implementations. The larger fragments of the CoreGram project will be documented in a series of book publications. The first book in this series was Müller, 2007b, which describes a fragment of German that is implemented in the grammar BerliGram. Three further books are in preparation and will be submitted to the series Implemented Grammars by Language Science Press: one on the Persian Grammar developed in the PerGram project (Müller, Samvelian and Bonami, In Preparation), the Danish Grammar developed in the DanGram project (Müller and Ørsnes, 2015) and the Mandarin Chinese grammar developed in the ChinGram project.

2 The Poverty of the Stimulus and Motivation of Analyses

Huge progress has been made in recent years in the area of language acquisition. Input-based methods with an utterance-final bias have been shown to explain acquisition data better than maturation explanations or principle and parameters models (Freudenthal et al., 2006, 2007, 2009). Bod (2009) showed that English auxiliary inversion can be learned even with the evidence that Chomsky (1971, p. 29–33) claimed to be necessary and unavailable. The cited research shows that quite elaborate structures can be learned from the input alone and hence if there is any innate language-specific knowledge at all it is probably rather general as assumed for instance by Hauser, Chomsky and Fitch (2002). The consequence for linguistic research is that the existence of certain structures in one language does not imply that such structures are part of the grammar of all languages. So, the existence of object agreement in Basque cannot be used as evidence for object agreement projections (AgrO) in German. Neither can the existence of postpositions and agreement in Hungarian be seen as evidence for AgrO projections and hidden movement processes in English. Such complicated analyses cannot be motivated language internally and hence are not acquirable from input alone.

Instead of imposing constraints from one language onto other languages, a bottom-up approach seems to be more appropriate: Grammars for individual languages should be motivated language internally. Grammars that share certain properties can be grouped in classes. This makes it possible to capture generalizations about groups of languages and language as such. Let us consider some examples: German, Dutch, Danish, English and French. If we start developing grammars for German and Dutch, we find that they share a lot of properties: both are SOV and V2 languages, both have a verbal complex. One main difference is the order of elements in the verbal complex. The situation can be depicted as in Figure 1 on the facing page. There are some properties that are shared between German and Dutch (Set 3). For instance, the argument structure, a list containing descriptions of syntactic and semantic properties of arguments, and the linking of these arguments to the meaning of the sign is contained in Set 3. In addition the constraints for SOV languages, the verb position in V2 clauses and the fronting of a constituent in V2 clauses are contained in Set 3. The respective constraints are shared between the two grammars. When we add another language, say Danish, we get further

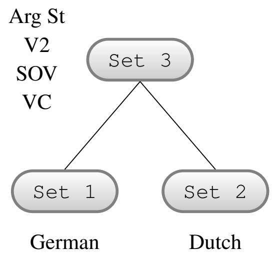

Figure 1. Shared properties of German and Dutch

differences. While German and Dutch are SOV, Danish is an SVO language. Figure 2 shows the resulting situation: The top-most node represents constraints that hold for all languages (for instance the argument structure constraints, linking and V2) and the node below it contains constraints that hold for German and Dutch only.[1] For instance these constraints contain constraints regarding verbal complexes and SOV order. The

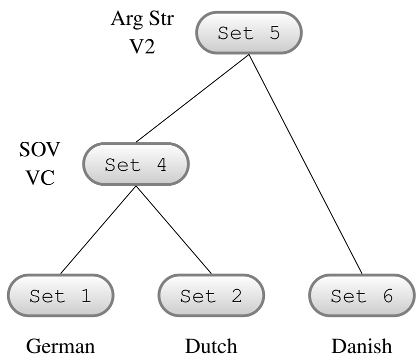

Figure 2. Shared Properties of German, Dutch, and Danish

union of Set 4 and Set 5 is the Set 3 of Figure 1.

If we add further languages further constraint sets will be distinguished. Figure 3 on the following page shows the situation that results when we add English and French. Again, the picture is not complete since there are constraints that are shared by Danish and English but not by French, but the general idea should be clear: by consequently

[1] In principle, there could be constraints that hold for Dutch and Danish but not for German and for German and Danish, but not for Dutch. These constraints would be removed from Set 1 and Set 2 respectively and put into another constraint set higher up in the hierarchy. These sets are not illustrated in the figure and we keep the names Set 1 and Set 2 for the constraint sets for German and Dutch.

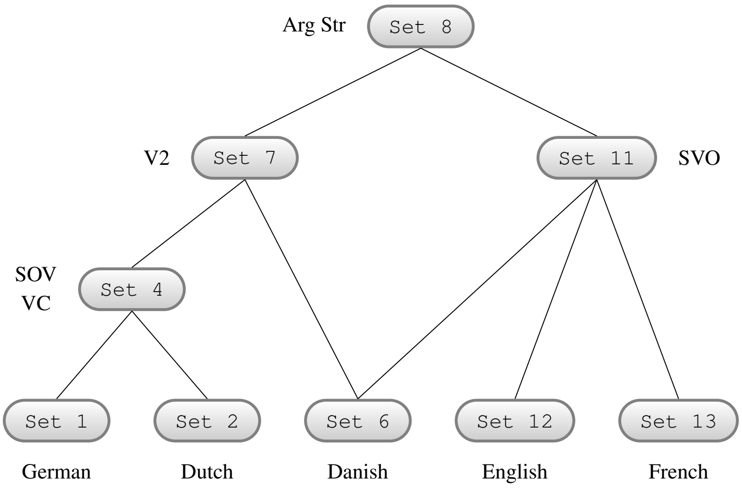

Figure 3. Languages and language classes

working this way, we should arrive at constraint sets that directly correspond to those that were established in the typological literature.

It should be clear from what has been said so far that the goal of every scientist that works this way is to find generalizations and to describe a new language in a way that reuses theoretical constructs that have been found useful for a language that is already covered. However, as was explained above the resulting grammars should be motivated by data of the respective languages and not by facts from other languages. In situations where more than one analysis would be compatible with a given dataset for language X the evidence from language Y with similar constructs is most welcome and can be used as evidence in favor of one of the two analyses for language X. I call this approach the *bottom-up approach with cheating*: unless there is contradicting evidence we can reuse analyses that have been developed for other languages.

3 Coverage and Highlights

The grammars of German, Persian, and Danish are relatively big. The German grammar (BerliGram) was the first one that was implemented. It is an extension of the grammars that were developed for the individual chapters of the HPSG text book (Müller, 2007b). The Situation Semantics was replaced by a Minimal Recursion Semantics (MRS, Copestake et al., 2005). MRS allows for underspecification of scope so that a sentence like (1a) gets one representation from which the several scopings can be derived.

(1) a. dass Max wieder alle Fenster öffnete
that Max again all windows opened
'that Max opened all windows again'

b. *again′*(∀(CAUSE(open))); repetitiv

c. *again′*(CAUSE(∀(open))); repetitiv

d. CAUSE(*again′*(∀(open))); restitutiv

The example in (1a) is an example of lexical subscoping: *öffnen* is lexically decomposed into CAUSE(open) and the *wieder* can scope below the CAUSE operator although there is no decomposition in syntax.

Apart from the modification of the semantics component further special phenomena were implemented. For instance an analysis of multiple frontings (Müller, 2003), something that is unique among the existing HPSG implementations. For a discussion of approaches to constituent order that are incompatible with the multiple fronting data see Müller, 2005, 2015b. Furthermore the analysis of depictives was added (Müller, 2008). Analyses that have been implemented in my earlier grammars of German have not yet been transferred to BerliGram.

The Danish grammar is documented in a 500+ page book, which is not complete yet. The following examples show in a compact way the interaction of several phenomena: passive with promotion of either the direct object or the indirect object (2a,c), passive and pronoun shift (2b,d), partial fronting and object shift in (2b,d):

(2) a. Bjarne bliver ikke anbefalet den.
Bjarne.NOM is not recommended it.ACC
'The book is not recommended to Bjarne.'

b. ? Anbefalet bliver Bjarne den ikke.
recommended is Bjarne.NOM it.ACC not
'The book is not recommended to Bjarne.'

c. Bogen bliver ikke anbefalet ham.
book.DEF.NOM is not recommended him.ACC
'The book is not recommended to him.'

d. ? Anbefalet bliver bogen ham ikke.
recommended is book.DEF.NOM him.ACC not
'The book is not recommended to him.'

The Mandarin Chinese grammar was implemented with the help of Jia Zhongheng. We used the description in Li and Thompson, 1981 as basis of our implementation. Among the things that are special are NPs that contain classifiers (3) and change of part of speech by reduplication as in (4).

(3)
na4 liang4 hong2 de che1 xiu4.le
that CL red DE car rust.ASP
'That red car rusts.'

The adjective gao1xing4 is converted into an adverb by forming the pattern AABB from the original adjective AB, that is gao1 is doubled and xing4 is doubled as well.

(4)
ta1 gao1gao1xing4xing4 you3yong3
he AABB=happily swims
'He swims happily.'

The Persian grammar is a larger fragment, which needs to be documented. The examples in (5) show lightverb constructions, which are an important feature of the language. (5a) shows that the future auxiliary can interrupt the preverb verb sequence of lightverbs. (5b) shows an example with the negation prefix and pro-drop.

(5) a. Man in kr r anjm xham dd.[2]
I this job DOM performance will-1SG gave
'I will do this work.'

b. mard râ dust nadât.
man DOM friend NEG-had
'he/she does not like the man.'

The Maltese grammar is an implementation of the description of Fabri, 1993. Fabri works in the framework of Lexical Decomposition Grammar, which is also a lexical framework and his analysis were translatable into HPSG without great efforts. The examples in (6) show definiteness marking. (6b) shows assimilation and (6c) shows clitics:

(6) a. Il-komunist xejjer lil-l-papa.
DEF-communist winks(3msg) Case-DEF-pope(msg)
'the communist winks the pope.'

b. It-terrorist bagat l-ittr-a lil-l-president.
DEF-terrorist sent DEF-letter-F Case-DEF-president
'The terrorist sent the president the letter.'

c. It-terrorist bagat=hie=l.
DEF-terrorist send.3M.SG=3F.SG=3M.SG
'The terrorist sent it to him.'

(6c) is ambiguous. There is a reading with clitic left dislocation. Both readings are found by the grammar.

4 Basic Assumptions

4.1 Valence

We assume that valence is represented in a uniform way across languages. Arguments of a head are represented in the ARG-ST list (Pollard and Sag, 1994, Chapter 9). They are mapped to the valence features SPR and COMPS in a language dependent fashion. For instance, English and Danish map the subject to the SPR list and the other arguments to COMPS. Danish inserts an expletive in cases in which there is no element that can be mapped to SPR, while English does not do this. German differs from both languages in mapping all arguments of finite verbs to the COMPS list.

The arguments in the ARG-ST list are ordered according to the obliqueness hierarchy of Keenan and Comrie (1977), which plays a role in the analysis of a variety of phenomena. The elements of the ARG-ST list are linked to the semantic roles that a certain head has to fill. Since the traditional role labels like agent and patient are problematic, we adopt Dowty's proto-role approach (Dowty, 1991). We use ARG1, ARG2, and so on as role labels.

[2] Karimi-Doostan, 1997, p. 73.

4.2 Constituent Structure and Constituent Order

Classical HPSG came with very view immediate dominance schemata: Head-Complement Schema, Head-Specifier Schema, Head-Adjunct Schema, the Head-Filler Schema for binding off unbounded dependencies, and the Head-Extra Schema for binding off clause bound nonlocal dependencies. Since Sag, 1997 many HPSG analyses have a more constructional flavor, that is, specific subconstructions of these general schemata are introduced (Sag, 2010). In the CoreGram project we stay within the old tradition of HPSG and keep the rather abstract dominance schemata. However, it is possible to state further constraints on the respective structures. So rather than having several very specific instances of the Head-Filler Schema, we have very few ones, for instance, for verb second clauses and relative clauses and formulate additional implicational constraints that constrain actual instances of head filler phrases further if the antecedent of the implicational constraint is true. Since the schemata are rather general they can be used for all languages under consideration so far. Of course the languages differ in terms of constituent order, but this can be dealt with by using linearization rules that are sensitive to features whose values are language specific. For instance, all heads have a feature INITIAL. The value is '+', if the head has to be serialized before its complements and '–' if it follows its complements. German and Persian verbs are INITIAL −, while English, Danish, Mandarin Chinese and Maltese verbs are INITIAL +.

We assume binary branching structures and hence we get the structures in (7) for English and the corresponding German example:

(7) a. He [[gave the woman] a book].

b. er [der Frau [ein Buch gab]]
he the woman a book gave

The LP rules enforce that *gave* is linearized before *the woman* and *gave the woman* is linearized before *a book*.

The scrambling of arguments is accounted for by ID schemata that allow the combination of a head with any of its arguments independent of the position an element has in the valence list of its head. Non-scrambling languages like English on the other hand combine heads with their complements in a strict order: the least oblique element is combined with the head first and then the more oblique complements follow. Non-scrambling languages with head-final order take the last element from the valence list first.

4.3 Morphology and Lexical Rules

We follow a lexical rule based approach to morphology. Lexical rules are basically unary branching trees that license new lexical items. A lexical rule can add or subtract to the phonology (in implementations the orthography) of an input item. For instance, it is possible to analyze the complex morphological patterns that we observe in Semitic languages by mapping a root consisting of consonants to a full-fledged stem or word that has the appropriate vowels inserted. We follow Bresnan and Mchombo (1995) in assuming the Lexical Integrity Principle. This means that all morphological combinations have to be done by lexical rules, that is, fully inflected forms are part of the lexicon, most of them being licensed by productive lexical rules.

Lexical rules do not have to change the phonology/orthography of the item they apply to. For instance lexical rules can be used to license further lexical items with extended or reduced valence requirements. As was argued in Müller, 2002, 2006 resultative constructions should be treated lexically. So there is a lexical rule that maps the stem *fisch-* of the intransitive version of the verb *fischen* ('to fish') onto a stem *fisch-* that selects for a secondary predicate (adjective or PP) and the subject of this predicate.

(8) Er fischt den Teich leer.
he fishes the pond empty

5 Implementation Details

5.1 TRALE

The grammars are implemented in TRALE. TRALE implements typed feature descriptions. Every grammar consists of a signature (a type hierarchy with feature introduction and appropriateness constraints) and a theory that states constraints on objects of these types. TRALE is implemented in Prolog and comes with an implementation of relational constraints that maps the TRALE relations to Prolog relations. TRALE has two parsers: a standard bottom-up chart parser and a linearization parser. The CoreGram project uses the standard bottom-up parser. Both parsers use a phrase structure backbone.

Compared to other systems like the LKB (Copestake, 2002) the expressive power of the description language is high (see also Melnik, 2007). This allows for the rather direct implementation of analyses that are proposed by theoretical linguists. The following descriptive devices are used in the theory and are provided by TRALE. The references point to papers who argue for such constructs.

- empty elements (Kiss, 1995; Meurers, 1999; Müller, 2007b; Levine and Hukari, 2006; Bender, 2000 und Sag, Wasow and Bender, 2003, p. 464; Borsley, 2004, Section 3.3; Müller, 2014b),
- relational constraints (Pollard and Sag, 1994; Bouma et al., 2001),
- complex antecedents in implicational constraints (Meurers, 2000, p. 207; Koenig and Davis, 2004, p. 145, 149; Müller, 2007b, p. 145, Section 10.3; Bildhauer and Cook 2010, p. 75),
- cyclic structures (Engdahl and Vallduví, 1994, p. 56, Meurers, 1999, Meurers, 2001, p. 176, Samvelian, 2007, p. 638), and
- a morphology component that has the expressive power that is needed to account for nontrivial morphological phenomena.

5.2 Setup of CoreGram

The grammars are organized in one directory for every language. The respective directories contain a subdirectory named Core-Grammar. This directory contains files that are shared between the grammars. For instance the file core-macros.pl contains macros that are or can be used by all languages. For every language there is a load file that loads the files from the core grammar directory that are relevant for the respective language. So, for instance english.pl, french.pl, and danish.pl all load nom-acc.pl since these languages are nominative-accusative languages. These files also contain code for loading macros and constraints for languages that do not form a verbal complex, while german.pl does load the files for cluster-forming languages. These files directly correspond to the constraint sets that were discussed in Section 2.

The possibility to specify type constraints makes it possible to specify constraints that hold for a certain construction cross-linguistically in a file that is loaded by all grammars and restrict structures of this type further in language particular files.

Lexical rules are also described by feature descriptions and organized in type hierarchies (Meurers, 2001). Like other constraints the constraints on lexical rules can be shared.

6 Measuring Progress

Much to the frustration of many linguists the contribution of certain approaches to progress in linguistics is rather unclear. Many proposals do not extend the amount of data that is covered in comparison to analyses that were developed during the 1980s in the framework of GB and other, non-transformational frameworks. In comparison the methodology described in Section 2 leads to grammars with increasing coverage and analyses that are improved by cross-linguistic considerations.

The TRALE system has been combined with [incr tsdb()], a software for systematic grammar profiling (Oepen and Flickinger, 1998). The grammars are accompanied with a set of example phrases that can be analyzed by the grammar. In addition the test suite files contain ungrammatical word sequences from the literature and ungrammatical sequences that were discovered during the grammar development process. Since TRALE has a chart display that makes it possible to inspect the parse chart, it is possible to inspect all linguistic objects that are licensed by the grammar, even if they do not play a role in analyzing the particular sentence under consideration. The result of this careful inspection is a collection of ungrammatical word sequences that no theoretical linguist would have been able to come up with since it is very difficult to find all the side effects that an analysis might have that is not constrained sufficiently. These negative examples are distributed with the grammars and book publications and can help theoretical and computational linguists improve their theories and implementations.

After changing a grammar the sentences of the respective test suite are parsed and the result can be compared to previous results. This way it is ensured that the coverage of grammars is extended. If constraints in files are changed that are shared with other grammars the respective grammars are tested as well. The effects that changes to grammar X cause in grammar Y are often unexpected and hence it is very important to do systematic testing.

7 Conclusions

We argued that linguistic theories have reached a level of complexity that cannot be handled by humans without help by computers. We discussed a method for constructing surface-oriented theories by extending the number of languages that are considered and finally provided a brief description of basic assumptions and the basic setup of the CoreGram project.

Bibliography

Bender, E. M. 2000. *Syntactic Variation and Linguistic Competence: The Case of AAVE Copula Absence*. Ph. D.thesis, Stanford University.

Bender, E. M. 2008. Grammar Engineering for Linguistic Hypothesis Testing. In N. G. et. al. (ed.), *Proceedings of the Texas Linguistics Society X Conference*, pp 16–36, Stanford CA: CSLI Publications ONLINE.

Bildhauer, F. 2008. *Representing Information Structure in an HPSG Grammar of Spanish*. Dissertation, Universität Bremen.

Bildhauer, F. and Cook, P. 2010. German Multiple Fronting and Expected Topic-Hood. In S. Müller (ed.), *Proceedings of the 17th International Conference on Head-Driven Phrase Structure Grammar, Université Paris Diderot*, pp 68–79, Stanford: CSLI Publications.

Bod, R. 2009. From Exemplar to Grammar: Integrating Analogy and Probability in Language Learning. *Cognitive Science* 33(4), 752–793.

Borsley, R. D. 2004. An Approach to English Comparative Correlatives. In Müller (2004), pp 70–92.

Bouma, G., Malouf, R. and Sag, I. A. 2001. Satisfying Constraints on Extraction and Adjunction. *Natural Language and Linguistic Theory* 19(1), 1–65.

Bresnan, J. and Mchombo, S. A. 1995. The Lexical Integrity Principle: Evidence from Bantu. *Natural Language and Linguistic Theory* 13, 181–254.

Chomsky, N. 1971. *Problems of Knowledge and Freedom*. London: Fontana.

Copestake, A. 2002. *Implementing Typed Feature Structure Grammars*. Stanford: CSLI Publications.

Copestake, A., Flickinger, D. P., Pollard, C. J. and Sag, I. A. 2005. Minimal Recursion Semantics: an Introduction. *Research on Language and Computation* 4(3), 281–332.

Dowty, D. R. 1991. Thematic Proto-Roles and Argument Selection. *Language* 67(3), 547–619.

Engdahl, E. and Vallduví, E. 1994. Information Packaging and Grammar Architecture: A Constraint-Based Approach. In E. Engdahl (ed.), *Integrating Information Structure into Constraint-Based and Categorial Approaches*, pp 39–79, Amsterdam: ILLC, DYANA-2 Report R.1.3.B.

Fabri, R. 1993. *Kongruenz und die Grammatik des Maltesischen*. Tübingen: Niemeyer.

Freudenthal, D., Pine, J. M., Aguado-Orea, J. and Gobet, F. 2007. Modeling the Developmental Patterning of Finiteness Marking in English, Dutch, German, and Spanish Using MOSAIC. *Cognitive Science* 31(2), 311–341.

Freudenthal, D., Pine, J. M. and Gobet, F. 2006. Modeling the Development of Children's Use of Optional Infinitives in Dutch and English Using MOSAIC. *Cognitive Science* 30(2), 277–310.

Freudenthal, D., Pine, J. M. and Gobet, F. 2009. Simulating the Referential Properties of Dutch, German, and English Root Infinitives in MOSAIC. *Language Learning and Development* 5(1), 1–29.

Hauser, M. D., Chomsky, N. and Fitch, W. T. 2002. The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science* 298, 1569–1579.

Karimi-Doostan, G. 1997. *Light Verb Constructions in Persian*. Ph. D.thesis, Department of Language and Linguistics, University of Essex.

Keenan, E. L. and Comrie, B. 1977. Noun Phrase Accessibility and Universal Grammar. *Linguistic Inquiry* 8(1), 63–99.

Kiss, T. 1995. *Infinite Komplementation. Neue Studien zum deutschen Verbum infinitum*. Tübingen: Niemeyer.

Koenig, J.-P. and Davis, A. R. 2004. Raising Doubts about Russian Impersonals. In Müller (2004).

Levine, R. D. and Hukari, T. E. 2006. *The Unity of Unbounded Dependency Constructions*. Stanford: CSLI Publications.

Li, C. N. and Thompson, S. A. 1981. *Mandarin Chinese. A Functional Reference Grammar*. Berkeley and Los Angeles: University of California Press.

Lipenkova, J. 2009. *Serienverbkonstruktionen im Chinesischen und ihre Analyse im Rahmen von HPSG*. Masters Thesis, Institut für Sinologie, Freie Universität Berlin.

Melnik, N. 2007. From "Hand-Written" to Computationally Implemented HPSG Theories. *Research on Language and Computation* 5(2), 199–236.

Meurers, D., Penn, G. and Richter, F. 2002. A Web-Based Instructional Platform for Constraint-Based Grammar Formalisms and Parsing. In D. Radev and C. Brew (eds.), *Effective Tools and Methodologies for Teaching NLP and CL*, pp 18–25.

Meurers, W. D. 1999. German Partial-VP Fronting Revisited. In G. Webelhuth, J.-P. Koenig and A. Kathol (eds.), *Lexical and Constructional Aspects of Linguistic Explanation*, pp 129–144, Stanford: CSLI Publications.

Meurers, W. D. 2000. Lexical Generalizations in the Syntax of German Non-Finite Constructions. Arbeitspapiere des SFB 340 No. 145, Eberhard-Karls-Universität, Tübingen.

Meurers, W. D. 2001. On Expressing Lexical Generalizations in HPSG. *Nordic Journal of Linguistics* 24(2), 161–217.

Müller, S. 1999. *Deutsche Syntax deklarativ. Head-Driven Phrase Structure Grammar für das Deutsche*. Tübingen: Niemeyer.

Müller, S. 2002. *Complex Predicates: Verbal Complexes, Resultative Constructions, and Particle Verbs in German*. Stanford: CSLI Publications.

Müller, S. 2003. Mehrfache Vorfeldbesetzung. *Deutsche Sprache* 31(1), 29–62.

Müller, S. (ed.). 2004. *Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar, Center for Computational Linguistics, Katholieke Universiteit Leuven*, Stanford, CSLI Publications.

Müller, S. 2005. Zur Analyse der deutschen Satzstruktur. *Linguistische Berichte* 201, 3–39.

Müller, S. 2006. Phrasal or Lexical Constructions? *Language* 82(4), 850–883.

Müller, S. 2007a. The Grammix CD Rom. A Software Collection for Developing Typed Feature Structure Grammars. In T. H. King and E. M. Bender (eds.), *Grammar Engineering across Frameworks 2007*, Stanford: CSLI Publications.

Müller, S. 2007b. *Head-Driven Phrase Structure Grammar: Eine Einführung*. Tübingen: Stauffenburg Verlag, first edition.

Müller, S. 2008. Depictive Secondary Predicates in German and English. In C. Schroeder, G. Hentschel and W. Boeder (eds.), *Secondary Predicates in Eastern European Languages and Beyond*, pp 255–273, Oldenburg: BIS-Verlag.

Müller, S. 2009a. A Head-Driven Phrase Structure Grammar for Maltese. In B. Comrie, R. Fabri, B. Hume, M. Mifsud, T. Stolz and M. Vanhove (eds.), *Introducing Maltese Linguistics*, pp 83–112, Amsterdam: John Benjamins Publishing Co.

Müller, S. 2009b. On Predication. In S. Müller (ed.), *Proceedings of the 16th International Conference on Head-Driven Phrase Structure Grammar, University of Göttingen, Germany*, pp 213–233, Stanford: CSLI Publications.

Müller, S. 2010. Persian Complex Predicates and the Limits of Inheritance-Based Analyses. *Journal of Linguistics* 46(3), 601–655.

Müller, S. 2012. On the Copula, Specificational Constructions and Type Shifting. Ms. Freie Universität Berlin.

Müller, S. 2014a. Artenvielfalt und Head-Driven Phrase Structure Grammar. In *Syntaxtheorien: Analysen im Vergleich*, pp 187–233, Tübingen: Stauffenburg.

Müller, S. 2014b. Elliptical Constructions, Multiple Frontings, and Surface-Based Syntax. In P. Monachesi, G. Jäger, G. Penn and S. Wintner (eds.), *Proceedings of Formal Grammar 2004, Nancy*, pp 91–109, Stanford: CSLI Publications.

Müller, S. 2015a. The CoreGram Project: Theoretical Linguistics, Theory Development and Verification To appear.

Müller, S. 2015b. *German Sentence Structure: An Analysis with Special Consideration of So-Called Multiple Fronting*. Berlin: Language Science Press, In Preparation.

Müller, S. and Ghayoomi, M. 2010. PerGram: A TRALE Implementation of an HPSG Fragment of Persian. In *Proceedings of 2010 IEEE International Multiconference on Computer Science and Information Technology – Computational Linguistics Applications (CLA'10). Wisła, Poland, 18–20 October 2010*, pp 461–467, Polnish Information Processing Society.

Müller, S. and Lipenkova, J. 2009. Serial Verb Constructions in Chinese: An HPSG Account. In S. Müller (ed.), *Proceedings of the 16th International Conference on Head-Driven Phrase Structure Grammar, University of Göttingen, Germany*, pp 234–254, Stanford: CSLI Publications.

Müller, S. and Ørsnes, B. 2011a. Positional Expletives in Danish, German, and Yiddish. In S. Müller (ed.), *Proceedings of the 18th International Conference on Head-Driven Phrase Structure Grammar, University of Washington, U.S.A.*, pp 167–187, Stanford: CSLI Publications.

Müller, S. and Ørsnes, B. 2011b. Positional Expletives in Danish, German, and Yiddish. In S. Müller (ed.), *Proceedings of the 18th International Conference on Head-Driven Phrase Structure Grammar*, pp 167–187, Stanford: CSLI Publications.

Müller, S. and Ørsnes, B. 2015. *Danish in Head-Driven Phrase Structure Grammar*. Berlin: Language Science Press, In Preparation.

Müller, S., Samvelian, P. and Bonami, O. In Preparation. *Persian in Head-Driven Phrase Structure Grammar*. Berlin: Language Science Press.

Oepen, S. and Flickinger, D. P. 1998. Towards Systematic Grammar Profiling. Test Suite Technology Ten Years After. *Journal of Computer Speech and Language* 12(4), 411–436, (Special Issue on Evaluation).

Ørsnes, B. 2009. Preposed Negation in Danish. In S. Müller (ed.), *Proceedings of the 16th International Conference on Head-Driven Phrase Structure Grammar, University of Göttingen, Germany*, pp 255–275, Stanford: CSLI Publications.

Penn, G. 2004. Balancing Clarity and Efficiency in Typed Feature Logic Through Delaying. In D. Scott (ed.), *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL'04), Main Volume*, pp 239–246, Barcelona, Spain.

Pollard, C. J. and Sag, I. A. 1994. *Head-Driven Phrase Structure Grammar*. Chicago, IL and London: University of Chicago Press.

Sag, I. A. 1997. English Relative Clause Constructions. *Journal of Linguistics* 33(2), 431–484.

Sag, I. A. 2010. English Filler-Gap Constructions. *Language* 86(3), 486–545.

Sag, I. A., Wasow, T. and Bender, E. M. 2003. *Syntactic Theory: A Formal Introduction*. Stanford: CSLI Publications, second edition.

Samvelian, P. 2007. A (Phrasal) Affix Analysis of the Persian Ezafe. *Journal of Linguistics* 43, 605–645.