Towards efficient HPSG generation for German, a non-configurational language

Berthold Crysmann[1] & Woodley Packard[2]

[1]CNRS, Laboratoire de linguistique formelle, Paris-Diderot [2]University of Washington

24th International Conference on Computational Linguistics (Coling 2012)
Mumbai, India, 10–14 December 2012

HPSG bottom-up generation and non-configurationality

- Techniques used to solve generation efficiency for English HPSG (ambiguity packing and index filtering; Carroll and Oepen, 2000) insufficient to tackle efficiency issues with GG, an HPSG for German
- Combinatorial explosion already for relatively short utterances
- Inefficiency of HPSG generation for German result of several conspiring factors, all relating to non-configurationality:
  - Relatively free word order (in clausal domain): late packing
    - permutation of complements
    - interspersal of modifiers
  - Verb placement
    - Alternation between V1/V2 and verb final clauses: bottom-up search must explore both left-branching and right-branching hypotheses
    - Initial placement gives rise to discontinuous lexical items (particle verbs)
  - Argument composition
    - Complex predicates form verb cluster, permitting interleaving of arguments
    - Verb movement may separate highest verb from cluster
  - Rich inflection

MRS term rewriting for generation efficiency

- Globally uninformed search with standard semantic input representation (MRS; Copestake et al., 2005) characterised by massive local ambiguity
- Large chunks of globally unsuccessful hypotheses can in principle be detected based on properties of the semantic input
- Information relevant for more informed search is only implicitly represented in the input
    - Information is out-of-place:
      ⇒ needs to be redundantly encoded locally on elementary predications
    - Information is represented by structural configurations in the input:
      ⇒ needs to be transformed into syntactic constraints
- Approach: Deterministically enrich semantic input with explicit syntacto-semantic constraints, suitable to early pruning of globally unsuccessful hypotheses

MRS enrichment

```
[ LTOP: h0
  INDEX: e19
  RELS: < [ prpstn_m_rel LBL: h0 MARG: h16  ARG0: e19 TPC: x17  PSV: u2 ]
   [ "_pron_n_ppro_rel" LBL: h3
     ARG0: x17 [ --TOP: + --COH: + --PUNCT: prop-punct --CAS: n-list PNG.PN: 2s ] ]
   [ "pronoun_q_rel" LBL: h7 ARG0: x17 RSTR: h9  body: h10  ]
   [ "_sollen_v_modal-haben_rel" LBL: h18
     ARG0: e19 [ --TOP: - --COH: - --SIND: 2s --PUNCT: prop-punct
                 --TPC: tpc-non-event-non-mod --SUB: -
                 TENSE: present MOOD: indicative PERFECTIVE: - ]
     ARG1: h14 ]
   [ "_schnarchen_v_n-haben_rel" LBL: h12
     ARG0: e15 [ --TOP: - --COH: - --PUNCT: prop-punct --TPC: tpc-non-event-non-mod
                 --SUB: bool TENSE: untensed PERFECTIVE: - ]
     ARG1: x17 ] >
  HCONS: < h18 qeq h16 h9 qeq h3 h14 qeq h12 > ]
```

Figure: Enriched input MRS for *Du sollst schnarchen* 'You should snore'

Enriching input semantics: Case

- Morphosyntactic case not represented in MRS input semantics (contrary to LFG f-structures)
- Baseline system therefore inflects every nominal expression (determiner, adjective, noun, pronoun) for all 4 possible cases
  Illicit case forms often detected very late during generation, because, with argument composition, case properties of inherited complements are locally unknown
- Case as a selected property can be derived from predicate-argument structure using linking information
- Case prediction must anticipate deviation from lexical case marking patterns
  - voice alternation (foregrounded role represented in semantics)
  - subject to object raising
  - conflicting case assignments in relativisation (exempted)
- Transfer grammar uses 35 rules to predict case of arguments

Enriching input semantics: Punctuation

- Punctuation in GG (Kilian, 2007) attached to words by inflectional rules
- Host word for sentence-final punctuation can only be determined after generation:
  (almost) every chart item inflected in 5 different ways (none, comma, period, exclamation and question mark)
- Sentence mode, however, represented in input MRS
- Distribute information about matrix sentence mode onto all EPs
- Restrict inflection rules (period, exclamation mark, question mark) to appropriate sentence mode

Enriching input semantics: Verb placement (SUB)

- German nonfinite verbs are final
- German finite verbs show placement alternation between initial (V1/V2) and final position
  - Final: relative clauses, embedded interrogatives, subclauses introduced by complementiser or subordinating conjunction
  - Initial: elsewhere
  - Initial/final word order can be detected from semantics, using
    - Sentence mode (declarative/interrogative)
    - Type of embedding (relativisation, complementation, conjunction)
    - presence of a topicalised element (embedded V2 vs. that-clause)
- Transfer grammar establishes for each verb "subordinate" vs. "non-subordinate" clause type
  - Predicts direction of branching for finite clauses
  - Predicts presence/absence of costly verb movement

Enriching input semantics: Complex predicates

- Auxiliaries, modals, raising and control verbs form clusters
- Arguments of predicates in a cluster can be interleaved

Letzte	*Woche*	*versprach*	*ein*	*Buch*	*er*	*ihm*	*zu*	*kaufen*
last	week	promised	a	book	he	him	to	buy

- Strong interaction with verb placement
  - Parser/Generator may encounter arguments of the initial verb, before knowing that verb's argument structure
  - Verb movement rules locally hypothesise additional dependents (subject and possibly object) whose identity is not known
  - Massive local ambiguity: ***any*** XP in the chart can satisfy hypothesised valencies

- MRS enrichment:
  - COH: restrict hypothesised valency to arguments of verb clusters
  - OIND: establish presence and identity of upstairs object
  - SIND: determine agreement for modals and raising verbs

Enriching input semantics: Extraction

- Long-distance dependencies are notorious source for inefficiency
- Extraction in German very common (including ordinary declaratives)
- External MRS marks topicalised element as information structural property of propositions and questions
- Gap prediction for argument extraction (TOP)
  - Redundantly mark the inherent variable of topicalised EP as [TOP +]
  - Mark all other inherent variables as [TOP −]
  - Exempt realitivised EPs ([TOP *na*]), as they are extracted within the relative clause, but may be in situ outside
  - Restrict extraction rules to *na_or_+* and head-valence rules to *na_or_−*
- Gap prediction for adjuncts (TPC)
  - Adjuncts not represented on argument structure of the modified head
  - Use semantic attachment information in the input to identify extraction site

Evaluation

- ACE generator
  - Open source runtime system for HPSG: parsing/generation/transfer
  - Implemented in C by Woodley Packard
  - Bottom-up chart generation similar to the LKB
  - Generation efficiency measures: ambiguity packing, index accessibility filtering (Carroll and Oepen, 2000)
  - Highly efficient: 14.7 times faster than LKB, 0.43s (vs. 6.34s) on Rondane treebank
- Test suites
  - 3 regression test suites for German (Babel, TSNLP, MRS)
    no test suite specifically designed for generation
    - Test items: 2259 total (MRS: 103, TSNLP: 1547, Babel: 609
    - Avg. string length: MRS: 4.44, TSNLP: 4.76, Babel: 6.76
  - Automatically parsed and manually disambiguated with Redwoods
  - Generation from gold standard MRS semantics
- Test parameters
  - All performance features tested in isolation and in combination

  +FEAT compare individual feature to baseline performance
  −FEAT: assess feature's effectiveness in combination (leave one out)

Main Results

- Combined effect of all efficiency measures yields significant performance improvement
  - speedup factor around 4.5 for MRS/TSNLP
  - speedup factor of 27.79 for Babel
- Space consumption reflects speedup, although savings are lower, owing to ambiguity packing (packed edges only counted once)
- Most features effective both in isolation and in combination

	MRS		TSNLP		Babel		MRS+TSNLP+Babel	
	Time (s)	Red.	Time (s)	Red.	Time (s)	Red.	Time (s)	Red.
Base	0.257	1.00	0.273	1.00	7.213	1.00	2.143	1.00
All	0.054	4.71	0.063	4.31	0.260	27.79	0.116	18.52

	MRS			TSNLP			Babel		
	Cov	Edges	Red.	Cov	Edges	Red.	Cov	Edges	Red.
Base	100.0	703	1.00	100.0	693	1.00	96.7	4864	1.00
All	100.0	116	6.07	100.0	172	4.03	100.0	554	8.78

Detailed results: Inflection features (*punct,cas*)

- Both inflection features consistently effective in isolation and combination
- PUNCT outperforms case prediction

  Explanations:
  - Punctuation prediction
    - is simple and exceptionless (5 rules)
    - applies to every part-of-speech (i.e. every lexical chart entry)
  - Case prediction
    - Complex and large rule set (35 rules)
    - Many cases of exemption (e.g. relatives)
    - only applies to nominal expressions

	MRS		TSNLP		Babel		MRS+TSNLP+Babel	
	Time (s)	Red.	Time (s)	Red.	Time (s)	Red.	Time (s)	Red.
Base	0.257	1.00	0.273	1.00	7.213	1.00	2.143	1.00
+CAS	0.222	1.16	0.216	1.26	4.547	1.59	1.384	1.55
+PUNCT	0.150	1.71	0.159	1.71	4.598	1.57	1.355	1.58
All	0.054	4.71	0.063	4.31	0.260	27.79	0.116	18.52
-CAS	0.056	4.59	0.065	4.20	0.371	19.46	0.147	14.58
-PUNCT	0.063	4.11	0.081	3.38	0.415	17.39	0.170	12.61

Detailed results: verb placement and argument composition

- Control for verb placement (SUB) beneficial throughout: controls both direction of branching and costly verb movement
- Control for argument composition and initial verb's argument (*oind*) more dependent on input complexity (when used in combination)
- *sind* (agreement on raising and modal verbs) far too item-specific

	MRS		TSNLP		Babel		MRS+TSNLP+Babel	
	Time (s)	Red.	Time (s)	Red.	Time (s)	Red.	Time (s)	Red.
Base	0.257	1.00	0.273	1.00	7.213	1.00	2.143	1.00
+SUB	0.210	1.22	0.215	1.27	6.042	1.19	1.786	1.20
+COH	0.226	1.14	0.275	0.99	5.681	1.27	1.730	1.24
+OIND	0.242	1.06	0.262	1.04	5.506	1.31	1.675	1.28
+SIND	0.262	0.98	0.274	1.00	7.035	1.03	2.096	1.02
All	0.054	4.71	0.063	4.31	0.260	27.79	0.116	18.52
-SUB	0.058	4.44	0.067	4.09	0.304	23.71	0.130	16.44
-COH	0.054	4.78	0.061	4.49	0.353	20.45	0.139	15.41
-OIND	0.056	4.59	0.062	4.37	0.526	13.72	0.187	11.46
-SIND	0.050	5.12	0.062	4.40	0.258	27.95	0.114	18.74

Detailed results: Extraction features (TOP / TPC)

- Extraction features consistently improve performance on all test suites
- Features for extracted arguments (TOP) and extraction site (TPC) complementary
  - TOP targets argument extraction, while TPC targets adjunct extraction
  - both features still contribute massively when used in combination
- TPC constitutes best overall single feature

	MRS		TSNLP		Babel		MRS+TSNLP+Babel	
	Time (s)	Red.	Time (s)	Red.	Time (s)	Red.	Time (s)	Red.
Base	0.257	1.00	0.273	1.00	7.213	1.00	2.143	1.00
+TOP	0.130	1.97	0.136	2.00	5.309	1.36	1.530	1.40
+TPC	0.131	1.96	0.178	1.53	3.602	2.00	1.099	1.95
All	0.054	4.71	0.063	4.31	0.260	27.79	0.116	18.52
-TOP	0.076	3.37	0.081	3.37	0.381	18.92	0.162	13.25
-TPC	0.064	4.01	0.079	3.46	0.897	8.04	0.299	7.17
-INDEX	0.049	5.24	0.069	3.97	0.433	16.67	0.166	12.91
-PACK	0.079	3.23	0.087	3.14	0.563	12.82	0.215	9.98

- Grammar-side efficiency features compare well to generator efficiency measures like packing and index filtering

Discussion: scalability

- Difference in strength of cumulative effect (factor 5.5) between less (MRS/TSNLP) vs. more (Babel) complex test suites
- Differences strongly correlate with input length
  Artificial reduction of string length of Babel to that of TSNLP (between 4.37 and 4.85) reduces difference to a factor between 1.2 and 2.1
  ⇒ Ceiling effect for less complex test suites
  ⇒ Method effectively contains exponential growth problem

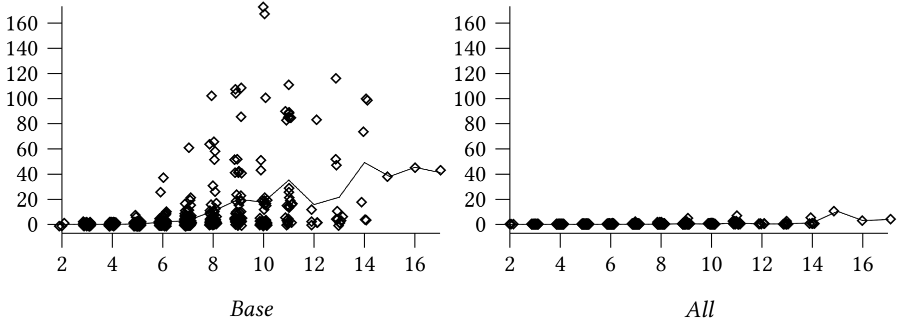

Figure: Processing time (in s) per string length (babel)

Discussion: scalability

- Difference in strength of cumulative effect (factor 5.5) between less (MRS/TSNLP) vs. more (Babel) complex test suites
- Differences strongly correlate with input length
  Artificial reduction of string length of Babel to that of TSNLP (between 4.37 and 4.85) reduces difference to a factor between 1.2 and 2.1
  ⇒ Ceiling effect for less complex test suites
  ⇒ Method effectively contains exponential growth problem

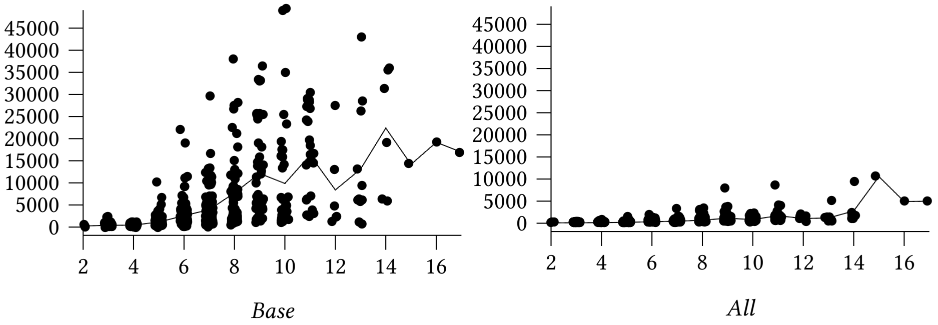

Figure: Passive edges per string length (babel)

Discussion: portability

- Deterministic enrichment of input semantics effective in closing performance gap between non-configurational German and configurational English
- Some features clearly language dependent: e.g., SUB and *oind*, which are restricted to V2 languages like German and Dutch
- Most features applicable to a wider range of languages
  - Extraction features
    Applicable to all languages with overt wh-movement or topicalisation
  - Case feature
    Applicable, inter alia, to Slavic languages
  - Punctuation feature
    - Most widely applicable
    - Tests with English Resource Grammar (Copestake and Flickinger, 2000) on Rondane test suite yield 10.5% reduction in processing time

Related work

- French SemFTAG (Gardent and Kow, 2007)
  - Enrichment of semantic generator input with tree features
  - Nearly deterministic paraphrase selection
- LFG (Zarrieß and Kuhn, 2010)
  - Transfer of syntactic f-structures to more general semantic representation
  - Inverse transfer step for argument structure to grammatical functions
    Massive difference in generator performance as to the detail of the rules (between 36.2s and 246.14s per item)
  - Major difference:
    Semantic input readily accepted by GG, enrichment is a pure performance measure
- EnJu HPSG (Nakanishi et al., 2005)
  - Beam search approach for generation efficency with tree bank-induced HPSG for English
  - Methods highly complementary:
    - Probabilistic pruning good at local decisions, e.g. for permutation of arguments
    - MRS-based enrichment permits non-local pruning with certainty
    - Combined system may yield good results at smaller beam sizes

Conclusion

- Rule-based method to improve generation efficiency with GG, an HPSG for German
  - Deterministic enrichment of semantic input with syntactic constraints
  - Grammar-side pruning on the basis of enriched input
- Evaluation on three different test suites demonstrates significant speedup factors
  - around a factor of 4.5 for relatively short utterances up to almost 28 for slightly more complex inputs
  - performance gains increase with input length
- Grammar-internal enrichment successfully attacks exponential growth problem while maintaining a general semantic interface for application development

ACE generator

- Overview
  - Runtime system for HPSG: parsing/generation/transfer
  - Implemented in C by Woodley Packard
  - Open source since 2011: `http://sweaglesw.org/linguistics/ace/`
  - Highly efficient: 14.7 times faster than LKB, 0.43s (vs. 6.34s) on Rondane treebank
- Bottom-up chart generation in ACE
  - Generation algorithm similar to the LKB, including efficiency measures like ambiguity packing and index accessibility filtering (Carroll and Oepen, 2000)
  - Reentrancies in the input made explicit by Skolem constants
  - Chart initialised with
    - lexical items whose semantics unifies with some elementary predication(s) in the input
    - semantically empty lexical items introduced by special trigger rules
  - Bottom-up creation of packed forest
  - Selective unpacking (using realisation model) and linearisation
  - Post-generation equivalence test (completeness)

Term rewrite system

- LKB and ACE feature an MRS term rewrite system, originally developed for semantic transfer in MT (Oepen et al., 2004, 2007)
- Transfer grammars are sequential resource-sensitive sets of rewrite rules
- Rewrite rules are tuples $< I, O, C, F >$ consisting of *Input*, *Output* and *Context* descriptions, as well as a *filter* condition:
  Replace input I with ouput O in context C, provided that the filter F does not match
- Each of the patterns in I, O, C and F may specify an arbitrary number of elementary predications
- Complexity and efficiency
  - Due to resource-sensitivity, order may be significant for multiple matches of the same rule, giving factorial worst-case complexity
  - MRS enrichment should not introduce ambiguity, so processing is deterministic, using first path only
  - Transfer costs are relatively small:
    Abs: 13.2ms (MRS); 12.1ms (TSNLP) and 29ms (Babel)
    Rel: 24.4% (MRS), 19.2% (TSNLP) and 11.3% (Babel) of overall processing time with full transfer grammar

Carroll, J. and Oepen, S. (2000). ‘High efficiency realization for a wide-coverage unification grammar’. In *Proceedings of the 2nd International Joint Conference on Natural Language Processing (IJCNLP)*. Jeju, Korea, 165–176.

Copestake, A. and Flickinger, D. (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.

Copestake, A., Flickinger, D., Pollard, C., and Sag, I. (2005). ‘Minimal recursion semantics: an introduction’. *Research on Language and Computation*, 3:281–332.

Gardent, C. and Kow, E. (2007). ‘A Symbolic Approach to Near-Deterministic Surface Realisation using Tree Adjoining Grammar’. In *45th Annual Meeting of the Association for Computational Linguistics - ACL 2007*. Prague, Czech Republic: Association for Computational Linguistics, 328–335.

Kilian, N. (2007). *Zum Punkt gekommen*. Master’s thesis, Universität des Saarlandes, Saarbrücken.

Nakanishi, H., Miyao, Y., and Tsujii, J. (2005). ‘Probabilistic models for disambiguation of an HPSG-based chart generator’. In *Proceedings of the Ninth International Workshop on Parsing Technology*. Association for Computational Linguistics, 93–102.

Oepen, S., Dyvik, H., Lønning, J. T., Velldal, E., Beermann, D., Carroll, J., Flickinger, D., Hellan, L., Johannessen, J. B., Meurer, P., Nordgård, T., and Rosén, V. (2004). ‘Som å kapp-ete med trollet? Towards MRS-based Norwegian–English Machine Translation’. In *Proceedings of the 10th International Conference on Theoretical and Methodological Issues in Machine Translation*. Baltimore, MD.

Oepen, S., Velldal, E., Lønning, J. T., Meurer, P., Rosén, V., and Flickinger, D. (2007). ‘Towards hybrid quality-oriented machine translation. On linguistics and probabilities in MT’. In *Proceedings of the 11th International Conference on Theoretical and Methodological Issues in Machine Translation*. Skövde, Sweden.

Zarrieß, S. and Kuhn, J. (2010). ‘Reversing f-structure rewriting for generation from meaning representations’. In M. Butt and T. H. King (eds.), *Proceedings of the LFG10 Conference, Ottawa*. Stanford: CSLI publications, 479–499.