

Re-Usable Tools for Precision Machine Translation*

Jan Tore Lønning[♣] and Stephan Oepen[♣]

[♣]Universitetet i Oslo, Computer Science Institute, Boks 1080 Blindern; 0316 Oslo (Norway)

[♣]Center for the Study of Language and Information, Stanford, CA 94305 (USA)

{jtl@ifi.uio.no|oe@csli.stanford.edu}

Abstract

The LOGON MT demonstrator assembles independently valuable general-purpose NLP components into a machine translation pipeline that capitalizes on output quality. The demonstrator embodies an interesting combination of hand-built, symbolic resources and stochastic processes.

1 Background

The LOGON projects aims at building an experimental machine translation system from Norwegian to English of texts in the domain of hiking in the wilderness (Oepen et al., 2004). It is funded within the Norwegian Research Council program for building national infrastructure for language technology (Fenstad et al., 2006). It is the goal for the program as well as for the project to include various areas of language technology as well as various methods, in particular symbolic and empirical methods. Besides, the project aims at reusing available resources and, in turn, producing re-usable technology.

In spite of significant progress in statistical approaches to machine translation, we doubt the long-term value of *pure* statistical (or data-driven) approaches, both practically and scientifically. To ensure grammaticality of outputs as well as felicity of the translation both linguistic grammars and deep semantic analysis are needed. The architecture of the LOGON system hence consists of a symbolic backbone system combined with various stochastic components for ranking system hypotheses. In a nutshell, a central research question in LOGON is to what degree state-of-the-art ‘deep’ NLP resources can contribute towards a precision MT system. We hope to engage the conference audience in some reflection on this question by means of the interactive presentation.

2 System Design

The backbone of the LOGON prototype implements a relatively conventional architecture, orga-

This demonstration reflects the work of a large group of people whose contributions we gratefully acknowledge. Please see ‘<http://www.emmtee.net>’ for background.

$$\langle h_1, \{ h_1:\text{proposition_m}(h_3), h_4:\text{proper_q}(x_5, h_6, h_7), h_8:\text{named}(x_5, \text{'Bodø'}), h_9:\text{populate_v}(e_2, \text{---}, x_5), h_9:\text{densely_r}(e_2) \}, \{ h_3 =_q h_9, h_6 =_q h_8 \} \rangle$$

Figure 1: Simplified MRS representation for the utterance ‘Bodø is densely populated.’ The core of the structure is a bag of *elementary predications* (EPs), using distinguished handles (‘ h_i ’ variables) and ‘ $=_q$ ’ (equal modulo quantifier insertion) constraints to underspecify scopal relations. Event- and instance-type variables (‘ e_j ’ and ‘ x_k ’, respectively) capture semantic linking among EPs, where we assume a small inventory of thematically bleached role labels (ARG₀ ... ARG _{n}). These are abbreviated through order-coding in the example above (see § 2 below for details).

nized around in-depth grammatical analysis in the source language (SL), semantic transfer of logical-form meaning representations from the source into the target language (TL), and full, grammar-based TL tactical generation.

Minimal Recursion Semantics The three core phases communicate in a uniform semantic interface language, Minimal Recursion Semantics (MRS; Copestake, Flickinger, Sag, & Pollard, 1999). Broadly speaking, MRS is a flat, event-based (neo-Davidsonian) framework for computational semantics. The abstraction from SL and TL surface properties enforced in our semantic transfer approach facilitates a novel combination of diverse grammatical frameworks, viz. LFG for Norwegian analysis and HPSG for English generation.

While an in-depth introduction to MRS (for MT) is beyond the scope of this project note, Figure 1 presents a simplified example semantics.

Norwegian Analysis Syntactic analysis of Norwegian is based on an existing LFG resource grammar, NorGram (Dyvik, 1999), under development on the Xerox Linguistic Environment (XLE) since around 1999. For use in LOGON, the grammar has been modified and extended, and it has been augmented with a module of Minimal Recursion Semantics representations which are computed from LFG f-structures by co-description.

In Norwegian, compounding is a productive morphological process, thus presenting the analysis engine with a steady supply of ‘new’ words, e.g. something like *klokkesletttrykk* meaning ap-

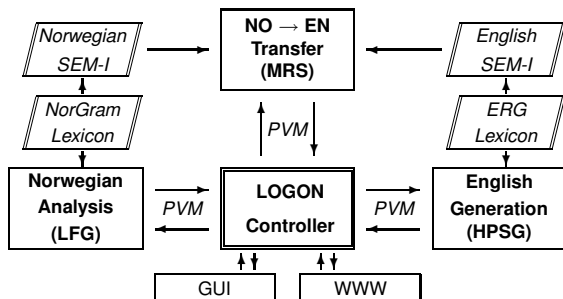


Figure 2: Schematic system architecture: the three core processing components are managed by a central controller that passes intermediate results (MRSs) through the translation pipeline. The Parallel Virtual Machine (PVM) layer facilitates distribution, parallelization, failure detection, and roll-over.

proximately *time-of-day expression*. The project uses its own morphological analyzer, compiled off a comprehensive computational lexicon of Norwegian, prior to syntactic analysis. One important feature of this processor is that it decomposes compounds in such a way that they can be compositionally translated downstream.

Current analysis coverage (including well-formed MRSS) on the LOGON corpus (see below) is approaching 80 per cent (of which 25 per cent are ‘fragmented’, i.e. approximative analyses).

Semantic Transfer Unlike in parsing and generation, there is less established common wisdom in terms of (semantic) transfer formalisms and algorithms. LOGON follows many of the main *VerbMobil* ideas—transfer as a resource-sensitive rewrite process, where rules replace MRS fragments (SL to TL) in a step-wise manner (Wahlster, 2000)—but adds two innovative elements to the transfer component, viz. (i) the use of typing for hierarchical organization of transfer rules and (ii) a chart-like treatment of transfer-level ambiguity. The general form of MRS transfer rules (MTRs) is as a quadruple:

$$[\text{CONTEXT:}] \text{INPUT} [\text{!FILTER}] \rightarrow \text{OUTPUT}$$

where each of the four components, in turn, is a partial MRS, i.e. triplet of a top handle, bag of EPs, and handle constraints. Left-hand side components are unified against an input MRS M and, when successful, trigger the rule application; elements of M matched by INPUT are replaced with the OUTPUT component, respecting all variable bindings established during unification. The optional CONTEXT and FILTER components serve to condition rule application (on the presence or absence of specific aspects of M), establish bindings for OUTPUT processing, but do *not* consume elements of M . Although our current focus is on

| Aggregate | total items # | word string ϕ | distinct trees ϕ | overall coverage % | time (s) ϕ |
|-------------------------------|---------------|--------------------|-----------------------|--------------------|-----------------|
| $30 < i\text{-length} < 40$ | 21 | 33.1 | 241.5 | 61.9 | 36.5 |
| $20 < i\text{-length} < 30$ | 174 | 23.0 | 158.6 | 80.5 | 15.7 |
| $10 < i\text{-length} < 20$ | 353 | 14.3 | 66.7 | 86.7 | 4.1 |
| $0 \leq i\text{-length} < 10$ | 495 | 4.6 | 6.0 | 90.1 | 0.7 |
| Total | 1044 | 11.6 | 53.50 | 86.7 | 4.3 |

(generated by [incr tsdb()] at 15-mar-2006 (15:51 h))

Table 1: Central measures of generator performance in relation to input ‘complexity’. The columns are, from left to right, the corpus sub-division by input length, total number of items, and average string length, ambiguity rate, grammatical coverage, and generation time, respectively.

translation into English, MTRs in principle state translational correspondence relations and, modulo context conditioning, can be reversed.

Transfer rules use a multiple-inheritance hierarchy with strong typing and appropriate feature constraints both for elements of MRSS and MTRs themselves. In close analogy to constraint-based grammar, typing facilitates generalizations over transfer regularities—hierarchies of predicates or common MTR configurations, for example—and aids development and debugging.

An important tool in the constructions of the transfer rules are the semantic interfaces (called SEM-Is, see below) of the respective grammars. While we believe that hand-crafted lexical transfer is a necessary component in precision-oriented MT, it is also a bottleneck for the development of the LOGON system, with its pre-existing source and target language grammars. We have therefore experimented with the acquisition of transfer rules by analogy from a bi-lingual dictionary, building on hand-built transfer rules as a seed set of templates (Nordgård, Nygaard, Lønning, & Oepen, 2006).

English Generation Realization of post-transfer MRSS in LOGON builds on the pre-existing LinGO English Resource Grammar (ERG; Flickinger, 2000) and LKB generator (Carroll, Copestake, Flickinger, & Poznanski, 1999). The ERG already produced MRS outputs with good coverage in several domains. In LOGON, it has been refined, adopted to the new domain, and semantic representations revised in light of cross-linguistic experiences from MT. Furthermore, chart generation efficiency and integration with stochastic realization have been substantially improved (Carroll & Oepen, 2005). Table 1 summarizes (exhaustive) generator performance on a segment of the LOGON

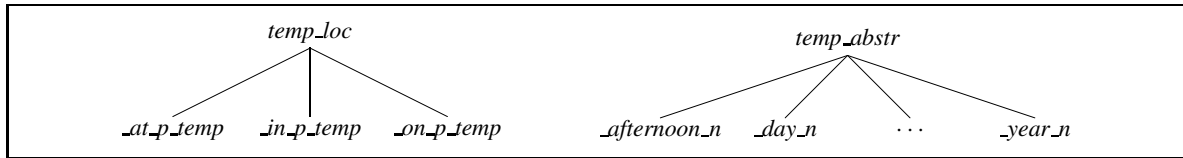


Figure 3: Excerpt from predicate hierarchies provided by English SEM-I. Temporal, directional, and other usages of prepositions give rise to distinct, but potentially related, semantic predicates. Likewise, the SEM-I incorporates some ontological information, e.g. a classification of temporal entities, though crucially only to the extent that is actually grammaticized in the language proper.

development corpus: realizations average at a little less than twelve words in length. After addition of domain-specific vocabulary and a small amount of fine-tuning, the ERG provides adequate analyses for close to ninety per cent of the LOGON reference translations. For about half the test cases, all outputs can be generated in less than one cpu second.

End-to-End Coverage The current LOGON system will only produce output(s) when all three processing phases succeed. For the LOGON target corpus (see below), this is presently the case in 35 per cent of cases. Averaging over actual outputs only, the system achieves a (respectable) BLEU score of 0.61; averaging over the entire corpus, i.e. counting inputs with processing errors as a zero contribution, the BLEU score drops to 0.21.

3 Stochastic Components

To deal with competing hypotheses at all processing levels, LOGON incorporates various stochastic processes for disambiguation. In the following, we present the ones that are best developed to date.

Training Material A corpus of some 50,000 words of edited, running Norwegian text was gathered and translated by three professional translators. Three quarters of the material are available for system development and also serve as training data for machine learning approaches. Using the discriminant-based Redwoods approach to treebanking (Oepen, Flickinger, Toutanova, & Manning, 2004), a first 5,000 English reference translations were hand-annotated and released to the public.¹ In on-going work on adapting the Redwoods approach to (Norwegian) LFG, we are working to treebank a sizable text segment (Rosén, Smedt, Dyvik, & Meurer, 2005; Oepen & Lønning, 2006).

Parse Selection The XLE analyzer includes support for stochastic parse selection models, assigning likelihood measures to competing analyses

¹See <http://www.delph-in.net/redwoods/> for the LinGO Redwoods treebank in its latest release, dubbed Norwegian Growth.

(Riezler et al., 2002). Using a trial LFG treebank for Norwegian (of less than 100 annotated sentences), we have adapted the tools for the current LOGON version and are now working to train on larger data sets and evaluate parse selection performance. Despite the very limited amount of training so far, the model already appears to pick up on plausible, albeit crude preferences (as regards topicalization, for example). Furthermore, to reduce fan-out in exhaustive processing, we collapse analyses that project equivalent MRSS, i.e. syntactic distinctions made in the grammar but not reflected in the semantics.

Realization Ranking At an average of more than fifty English realizations per input MRS (see Table 1), ranking generator outputs is a vital part of the LOGON pipeline. Based on a notion of automatically derived *symmetric treebanks*, we have trained comprehensive discriminative, log-linear models that (within the LOGON domain) achieve up to 75 per cent exact match accuracy in picking the most likely realization among competing outputs (Vellidal & Oepen, 2005). The best-performing models make use of configurational (in terms of tree topology) as well as of string-level properties (including local word order and constituent weight), both with varied domains of locality. In total, there are around 300,000 features with non-trivial distribution, and we combine the MaxEnt model with a traditional language model trained on a much larger corpus (the BNC). The latter, more standard approach to realization ranking, when used in isolation only achieves around 50 per cent accuracy, however.

4 Implementation

Figure 2 presents the main components of the LOGON prototype, where all component communication is in terms of sets of MRSS and, thus, can easily be managed in a distributed and (potentially) parallel client-server set-up. Both the analysis and generation grammars ‘publish’ their interface to transfer—i.e. the inventory and synopsis of seman-

tic predicates—in the form of a Semantic Interface specification (‘SEM-I’; Flickinger, Lønning, Dyvik, Oepen, & Bond, 2005), such that transfer can operate without knowledge about grammar internals. In practical terms, SEM-Is are an important development tool (facilitating well-formedness testing of interface representations at all levels), but they also have interesting theoretical status with regard to transfer. The SEM-Is for the Norwegian analysis and English generation grammars, respectively, provide an exhaustive enumeration of legitimate semantic predicates (i.e. the transfer vocabulary) and ‘terms of use’, i.e. for each predicate its set of appropriate roles, corresponding value constraints, and indication of (semantic) optionality of roles. Furthermore, the SEM-I provides generalizations over classes of predicates—e.g. hierarchical relations like those depicted in Figure 3 below—that play an important role in the organization of MRS transfer rules.

5 Open-Source Machine Translation

Despite the recognized need for translation, there is no widely used open-source machine translation system. One of the major reasons for this lack of success is the complexity of the task. By association to the international open-source DELPH-IN effort² and with its strong emphasis on re-usability, LOGON aims to help build a repository of open-source precision tools. This means that work on the MT system benefits other projects, and work on other projects can improve the MT system (where EBMT and SMT systems provide results that are harder to re-use). While the XLE software used for Norwegian analysis remains proprietary, we have built an open-source bi-directional Japanese–English prototype adaptation of the LOGON system (Bond, Oepen, Siegel, Copestake, & Flickinger, 2005). This system will be available for public download by the summer of 2006.

References

Bond, F., Oepen, S., Siegel, M., Copestake, A., & Flickinger, D. (2005). Open source machine translation with DELPH-IN. In *Proceedings of the Open-Source Machine Translation workshop at the 10th Machine Translation Summit* (pp. 15–22). Phuket, Thailand.

Carroll, J., Copestake, A., Flickinger, D., & Poznanski, V. (1999). An efficient chart generator for (semi-)lexicalist grammars. In *Proceedings of the 7th European Workshop on Natural Language Generation* (pp. 86–95). Toulouse, France.

²See ‘<http://www.delph-in.net>’ for details, including the lists of participating sites and already available resources.

Carroll, J., & Oepen, S. (2005). High-efficiency realization for a wide-coverage unification grammar. In R. Dale & K.-F. Wong (Eds.), *Proceedings of the 2nd International Joint Conference on Natural Language Processing* (Vol. 3651, pp. 165–176). Jeju, Korea: Springer.

Copestake, A., Flickinger, D., Sag, I. A., & Pollard, C. (1999). *Minimal Recursion Semantics. An introduction*. In preparation, CSLI Stanford, Stanford, CA.

Dyvik, H. (1999). The universality of f-structure. Discovery or stipulation? The case of modals. In *Proceedings of the 4th International Lexical Functional Grammar Conference*. Manchester, UK.

Fenstad, J.-E., Ahrenberg, L., Kvale, K., Maegaard, B., Mühlenbock, K., & Heid, B.-E. (2006). KUNSTI. Knowledge generation for Norwegian language technology. In *Proceedings of the 5th International Conference on Language Resources and Evaluation*. Genoa, Italy.

Flickinger, D. (2000). On building a more efficient grammar by exploiting types. *Natural Language Engineering*, 6(1), 15–28.

Flickinger, D., Lønning, J. T., Dyvik, H., Oepen, S., & Bond, F. (2005). SEM-I rational MT. Enriching deep grammars with a semantic interface for scalable machine translation. In *Proceedings of the 10th Machine Translation Summit* (pp. 165–172). Phuket, Thailand.

Nordgård, T., Nygaard, L., Lønning, J. T., & Oepen, S. (2006). Using a bi-lingual dictionary in lexical transfer. In *Proceedings of the 11th conference of the European Association of Machine Translation*. Oslo, Norway.

Oepen, S., Dyvik, H., Lønning, J. T., Velldal, E., Beermann, D., Carroll, J., Flickinger, D., Hellan, L., Johannessen, J. B., Meurer, P., Nordgård, T., & 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., Flickinger, D., Toutanova, K., & Manning, C. D. (2004). LinGO Redwoods. A rich and dynamic treebank for HPSG. *Journal of Research on Language and Computation*, 2(4), 575–596.

Oepen, S., & Lønning, J. T. (2006). Discriminant-based MRS banking. In *Proceedings of the 5th International Conference on Language Resources and Evaluation*. Genoa, Italy.

Riezler, S., King, T. H., Kaplan, R. M., Crouch, R., Maxwell, J. T., & Johnson, M. (2002). Parsing the Wall Street Journal using a Lexical-Functional Grammar and discriminative estimation techniques. In *Proceedings of the 40th Meeting of the Association for Computational Linguistics*. Philadelphia, PA.

Rosén, V., Smedt, K. D., Dyvik, H., & Meurer, P. (2005). TrePil. Developing methods and tools for multilevel treebank construction. In *Proceedings of the 4th Workshop on Treebanks and Linguistic Theories* (pp. 161–172). Barcelona, Spain.

Velldal, E., & Oepen, S. (2005). Maximum entropy models for realization ranking. In *Proceedings of the 10th Machine Translation Summit* (pp. 109–116). Phuket, Thailand.

Wahlster, W. (Ed.). (2000). *Verbmobil. Foundations of speech-to-speech translation*. Berlin, Germany: Springer.