Som å kapp-ete med trollet?[*]

— Towards MRS-Based Norwegian – English Machine Translation —

Stephan Oepen[♣♠], Helge Dyvik[◇], Jan Tore Lønning[♣], Erik Velldal[♣],
Dorothee Beermann[♡], John Carroll[△], Dan Flickinger[♠],
Lars Hellan[♡], Janne Bondi Johannessen[♣], Paul Meurer[◇],
Torbjørn Nordgård[♡], and Victoria Rosén[◇]

[♣]Universitetet i Oslo, Boks 1102 Blindern; 0317 Oslo (Norway)
[◇]Universitetet i Bergen, Sydnesplassen 7, 5007 Bergen (Norway)
[♡]Norges Teknisk-Naturvitenskapelige Universitet, 7491 Trondheim (Norway)
[♠]Center for the Study of Language and Information, Stanford, CA 94305 (USA)
[△]University of Sussex, Falmer, Brighton BN1 9QH (UK)

`all@emmtee.net`

Abstract

We present a relatively large-scale initiative in high-quality MT based on semantic transfer, reviewing the motivation for this approach, general architecture and components involved, and preliminary experience from a first round of system integration (to be accompanied by a hands-on system demonstration, if appropriate).

> *The translation problem is one whose solution must be searched incrementally.*
> *There will be no dramatic event to signal the end of the search.*
> (Martin Kay, 1982)

1 Introduction

About four years after the completion of the giant Verb*mobil* project (Wahlster, 2000), we have seen little new work on machine translation (MT) research grounded in precise linguistic theory. We present the Norwegian national initiative LOGON—a publicly funded effort twice the scope of Verb*mobil*, at least when viewed in proportion to the populations of Germany and Norway—which aims to deliver high-quality, document-level Norwegian – English MT based on the combination of a symbolic, semantic-transfer-oriented backbone and stochastic processes for ambiguity management and robustness.

The current prototype targets tourism-related publications and, albeit very limited in coverage to date, instantiates a novel set-up: the use of Minimal Recursion Semantics (MRS; Copestake, Flickinger, Sag, & Pollard, 2003) in an MT system that employs different grammatical frameworks for analysis and generation (viz. LFG and HPSG, respectively). The following sections motivate our 'deep' approach to MT (§ 2), sketch the LOGON system architecture and resources involved (§ 3), and then present novel aspects of individual components briefly (§ 4 to § 6). We conclude with a few remarks on accompanying LOGON activities and short- and mid-term research plans (§ 7 and § 8, respectively).

---

[*] Norwegian idiom; literally: 'Like contest-eating with the troll'. According to a traditional Norwegian tale, young Askeladden got engaged in an eating match with a troll, and tricked him by spooning food into a bag he had hidden under his shirt. When the boy eventually cut open the bag, ostensibly to make more room, the ogre imitated him and got killed by his own hand.

$$
\begin{array}{l}
\langle\, h_1, \\
\quad \{\, h_1{:}\textsf{proposition\_m}(h_3),\ h_4{:}\textsf{proper\_q}(x_5,\ h_6,\ h_7),\ h_8{:}\textsf{named}(x_5, \text{`Bodø'}), \\
\qquad h_9{:}\textsf{\_populate\_v}(e_2,\ \_,\ x_5),\ h_9{:}\textsf{\_densely\_r}(e_2)\,\}, \\
\quad \{\, h_3 =_q h_9,\ h_6 =_q h_8 \,\}\,\rangle
\end{array}
$$

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) to express scopal relations, where handle identity denotes scopal conjunction and an additional set of '$=_q$' (equal modulo quantifier insertion) handle constraints enables scope underspecification. Event- and instance-type variables ('e_j' and 'x_k', respectively) capture semantic linking among EPs, where MRSs tend to use a small inventory of thematically bleached role labels (ARG_0 ... ARG_n), abbreviated through order-coding in the example above (see § 3 below for details).

2 Why Deep MT in the 21$^{\text{st}}$ Century?

Even if we stay away from idioms involving Norwegian folk tales, machine translation is a hard, if not impossible, problem. The task encompasses not only all strata of linguistic description—phonology to discourse—but in the general case requires potentially unlimited knowledge about the actual world and situated language use. The state-of-the-art in MT today is somewhat ambivalent: most commercial systems deploy rule-based approaches (often using techniques from the 1960s and 1970s) while MT research has been dominated by statistical methods for the past decade or so.

Like a growing number of colleagues, we doubt the long-term value of *pure* statistical (or data-driven) approaches, both practically and scientifically. Large parallel training corpora are scarce for most languages, and word-level alignment remains a challenging research topic. Assuming sufficient training material, statistical translation quality still leaves much to be desired, and probabilistic NLP experience in general suggests that we can expect 'ceiling' effects on system evolution. Statistical MT research has yet to find a satisfactory role for linguistic analysis, albeit shallow; by itself, it does not further our understanding of language.

Emerging work on combining symbolic and data-driven approaches to MT will depend on an active stream of state-of-the-art, MT-oriented linguistics research. LOGON capitalizes on linguistic precision for high-quality translation and, accordingly, puts scalable linguistic resources—complemented with stochastic components—at the core of the initiative. Despite frequent cycles of overly high hopes and subsequent disillusionment, MT in our view is the type of application that may demand knowledge-heavy, 'deep' approaches to NLP for its ultimate, long-term success. Plurality of approaches to grammatical description and reusability of individual parts are among the strong points of the LOGON collaboration.

3 System Architecture & Interface Representations

Reflecting the above desiderata (as well as existing expertise and research interests among partners), the LOGON prototype implements a relatively conventional architecture, organized 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. The three core phases communicate in a uniform semantic interface language, Minimal Recursion Semantics (MRS). Although MRS was originally proposed for, among others, MT tasks (Copestake, Flickinger, Malouf,

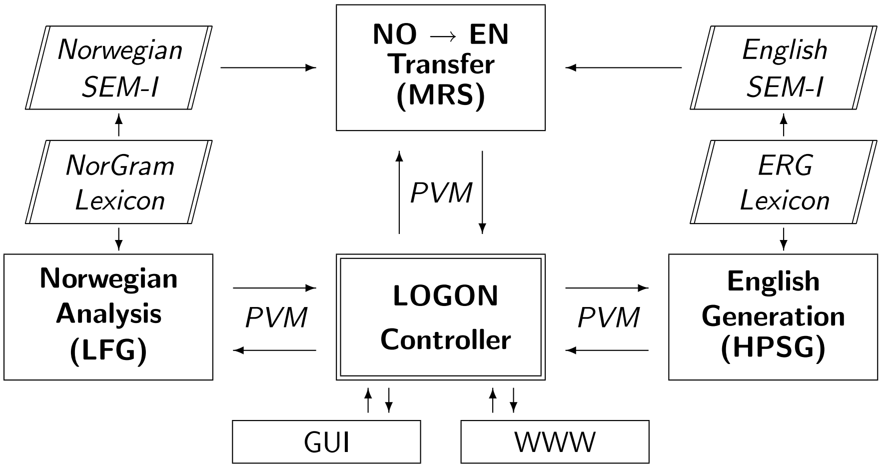

Figure 2: Schematic system architecture: the three core processing components are managed by a central controller that passes intermediate results (MRSs) through the translation pipeline. The Parallel Virtual Machine (PVM) layer facilitates distribution, parallelization, failure detection, and roll-over.

Riehemann, & Sag, 1995), to our best knowledge LOGON is the first system to implement end-to-end MRS-based translation. Broadly speaking, MRS is a flat, event-based (neo-Davidsonian) framework for computational semantics that facilitates underspecification of both scope relations[1] and generalization over classes of predicates (e.g. two-place temporal relations corresponding to distinct lexical prepositions: 'in May' vs. 'on Monday'; see § 5 below), thus enabling MT components to defer the resolution of ambiguity and supporting flexible experimentation with 'moving' distinctions around between the analysis, transfer, and generation phases. Furthermore, 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 note (but see references cited above), Figure 1 presents an example semantics for a simplified example from the LOGON corpus. The truth-conditional core is captured as a flat multiset (or 'bag') of *elementary predications* (EPs), combined with generalized quantifiers and designated *handle* variables to account for scopal relations. The bag of EPs, often termed the `RELS` set, is complemented by the handle of the top-scoping EP and a set of *handle constraints* (or `HCONS`) recording restrictions on scope relations contributed by the syntax.

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 semantic predicates—in the form of a Semantic Interface specification ('SEM-I'), such that transfer can operate without knowledge about grammar internals. In practical terms SEM-Is are an important

---

[1] In this respect, MRS is closely related to a tradition of underspecified semantics reflected in, among others, Quasi-Logical Form (QLF; Alshawi & Crouch, 1992), Underspecified Discourse Representation Theory (UDRT; Reyle, 1993), Hole Semantics (Bos, 1995), and the Constraint Language for Lambda Structures (CLLS; Egg, Koller, & Niehren, 2001).

development tool (facilitating wellformedness 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 4 below—that play an important role in the organization of MRS transfer rules.

4 Norwegian Analysis Using LFG

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. The grammar has a lexicon comprising some 80,000 lemmas. For use in LOGON, the grammar is extended and augmented with an MRS module, deriving representations suitable as input to transfer. NorGram assigns the usual LFG representations c-structure (PS tree) and f-structure (attribute-value matrix expressing, *inter alia*, grammatical relations like subject and object) to sentences. The f-structure is derived by co-description: partial descriptions of f-structures are associated with c-structure rules and lexical entries. The LFG architecture allows the projection of new representations by similar co-description, and the MRS-structure is projected off the f-structure in this way. Thus, the projection architecture of LFG allows the derivation of representations meeting external specifications, which is a desirable property of resource grammars.

We may consider a simplified example. In each equation, let '↑' denote the f-structure projected by the mother c-structure node (as usual in LFG), and let 'm::f' denote the MRS structure projected by the f-structure f. Furthermore, let '`KEY`' be an attribute whose value is to be one elementary predication, i.e. the kind of basic structures which are collected in a set labeled '`RELS`' in the flat MRS representations, and let '`p`' be the predicate introduced by a verb. Simplifying some details, the following path equations are among those associated with transitive verbs in the NorGram lexicon:

(1) $(\uparrow\texttt{PRED}) = \text{'}\texttt{p}\langle(\uparrow\texttt{SUBJ})(\uparrow\texttt{OBJ})\rangle\text{'}$
(2) $(\texttt{m::}\uparrow\texttt{KEY RELATION}) = \text{'}\texttt{p}\text{'}$
(3) $(\texttt{m::}\uparrow\texttt{KEY ARG1}) = (\texttt{m::}(\uparrow\texttt{SUBJ})\ \texttt{KEY ARG0})$
(4) $(\texttt{m::}\uparrow\texttt{KEY ARG2}) = (\texttt{m::}(\uparrow\texttt{OBJ})\ \texttt{KEY ARG0})$
(5) $(\texttt{m::}\uparrow\texttt{KEY}) \in (\texttt{m::}\uparrow\texttt{RELS})$
(6) $(\texttt{m::}\uparrow\texttt{RELS}) = (\texttt{m::}(\uparrow\texttt{SUBJ})\ \texttt{RELS})$
(7) $(\texttt{m::}\uparrow\texttt{HCONS}) = (\texttt{m::}(\uparrow\texttt{SUBJ})\ \texttt{HCONS})$
(8) $(\texttt{m::}\uparrow\texttt{RELS}) = (\texttt{m::}(\uparrow\texttt{OBJ})\ \texttt{RELS})$
(9) $(\texttt{m::}\uparrow\texttt{HCONS}) = (\texttt{m::}(\uparrow\texttt{OBJ})\ \texttt{HCONS})$

Equation (3) unifies the `ARG1` of the verb's `KEY` with the `ARG0` of the subject's `KEY`, and equation (4) similarly links `ARG2` with the object's `ARG0`. The symbol '=' expresses union when its arguments are sets. Thus, equations (6) – (9) take the unions of the sets of EPs (`RELS`) and handle constraints (`HCONS`) of the various constituents, thus creating the bags

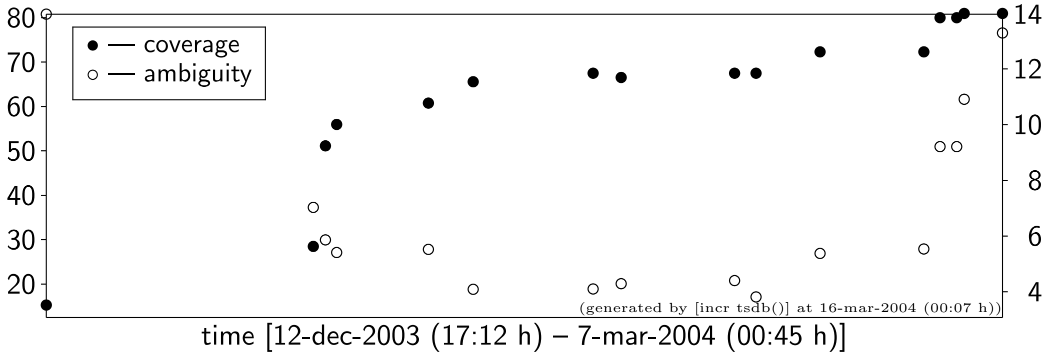

Figure 3: Evolution of analysis coverage (in per cent, left axis) and average structural ambiguity (number of analyses, right axis) over a three-month development interval.

of such elements characteristic of MRS. The MRS projection of NorGram presently contains about 2000 equations of this kind.

A limited amount of post-processing is applied to the MRS-structures produced by the grammar to arrive at the appropriate representation format for transfer, e.g. construction of the appropriate relation names and co-indexed variables. Furthermore, sometimes the desired scope and other hierarchical relations in the MRS representation may deviate from the descriptions in the resource grammar. Unless the analysis in the resource grammar is changed in such cases, it may be inconvenient to derive the appropriate MRS analysis by simple co-description. One example is the analysis of adverbial clauses, in which the relations introduced by subjunctions like *if*, *when*, *because*, etc. are analyzed as the top-level predicates of the top proposition in the MRS structure, taking the main clause and subordinate clause propositions as scopal arguments, while NorGram—more traditionally—has such subjunctions subordinate to the main clause. In a few such cases the grammar-derived MRS structures are post-processed slightly to derive the appropriate representations.

While the c- and f-structures jointly define the set of grammatical strings in the language, the MRS projection does not constrain grammaticality, but rather exploits f-structure information to assemble representations in the MRS format. Since the MRS projection of NorGram exploits the information in the f-structure, it could serve as a basis for developing similar MRS projections for parallel LFG grammars for other languages, such as the grammars developed in the ParGram project (see Dyvik, 2003, for a brief ParGram overview).

Figure 3 plots the development over a three-month integration period in NorGram coverage and average structural ambiguity with respect to an initial development corpus comprised of excerpts from tourism brochures. The initial jumps in coverage are due to the addition of domain-specific compounds and proper names to the lexicon. About twenty per cent of the sentences covered to date have fragmented analyses, resulting from chunk parsing triggered when no complete analysis is found. XLE fragment parsing makes use of optimality markings to select analyses with maximal-sized fragments. Our goal is to transfer and generate from fragmented MRSs as well, which will ensure some translation results even for sentences that do not parse completely. The XLE analyzer includes support for stochastic parse selection models, assigning likelihood measures to competing analyses, and we are in the process of preparing Norwegian training data.

5 Towards MRS-Based Transfer

Unlike in parsing and generation frameworks, there is less established common wisdom in terms of (semantic) transfer formalisms and algorithms. LOGON follows many of the main Verb*mobil* 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

[CONTEXT :] INPUT [! FILTER] → 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 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 (the LKB formalism; Copestake, 2002) 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. The following is a simplified example, using the '$\boxed{i}$' notation for variable binding, '_' for underspecification, and omitting the MRS top handle and handle constraints where appropriate:

$\mathsf{arg12\_mtr} := \langle \boxed{h_0}, \{ \boxed{h_1}{:}\_(\boxed{e_2}, \boxed{i_3}, \boxed{x_4}) \}, \{\} \rangle \rightarrow \langle \boxed{h_0}, \{ \boxed{h_1}{:}\_(\boxed{e_2}, \boxed{i_3}, \boxed{x_4}) \}, \{\} \rangle$

$\mathsf{arg12\_mtr} \wedge \{ \mathsf{befolke\_v} \} \xrightarrow{?} \{ \mathsf{populate\_v} \}$

$\mathsf{arg12\_mtr} \wedge \{ \mathsf{befolke\_v} \} \rightarrow \{ \mathsf{inhabit\_v} \}$

Both translations of the Norwegian *befolke* instantiate the same MTR type, inheriting all its properties, i.e. the parallelism of all handle, event, and instance variables. Transfer rules are organized as a sequence of ordered rule sets, where the overall ordering of MTR sets and ordering of rules in each set is supplied by the transfer grammar.[2] The resource-sensitive nature of the rewrite process is due to each rule application consuming parts of the MRS at that stage, thus changing the environment in which subsequent rule applications will be evaluated. In the example above, the first of the two MTRs is marked optional (the '$\xrightarrow{?}$'operator), while the second is obligatory: applying transfer rules in this order will

---

[2] The provision for successive application of several MTR sets is mostly for convenience, enabling transfer grammars to organize the rewrite process into multiple phases and optionally apply output filters upon the completion of each phase. The LOGON transfer grammar, for example, includes two sets of language-specific MTRs to accommodate grammar-specific idiosyncrasies before and after the core transfer phase, in some cases simply suppressing superfluous information (e.g. predicates introduced by selected-for prepositions and some aspectual markers), in others re-arranging or augmenting semantics to facilitate English generation. Transfer outputs incorporating plural mass nouns, for example, require the insertion of a suitable 'classifier' (Bond, Ogura, & Ikehara, 1996), in order to generate, say, *two pieces of information* instead of the ungrammatical **two informations.*

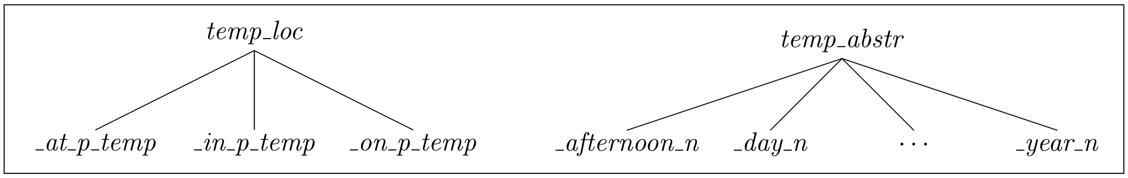

Figure 4: Excerpt from predicate hierarchies provided by English SEM-I. Temporal, directional, and other usages of prepositions give rise to distinct, but potentially related, semantic predicates. Likewise, the SEM-I incorporates some ontological information, e.g. a classification of temporal entities, though crucially only to the extent that is actually grammaticized in the language proper.

create a non-deterministic fork in the rewrite process and gives rise to multiple outputs. Making the last MTR obligatory ensures that SL 'befolke_v', ultimately, is transferred into a TL predicate (and at the same time significantly cuts down on the transfer search space). As each transfer rule may have to apply to its own output (e.g. in the face of MRSs with multiple occurrences of the same predicate), the LOGON transfer component implements chart-like factoring of ambiguity under MRS 'equivalence' (common normal form) to avoid combinatorial explosion and spurious permutations.[3]

To further illustrate MRS-based transfer, consider the following example:

n_n_lexicalization_mtr :=
⟨$\boxed{h_0}$,{$\boxed{h_1}$:_($\boxed{x_2}$), $\boxed{h_3}$:_($\boxed{x_4}$), $\boxed{h_3}$:unspec($\boxed{x_4}$, $\boxed{x_2}$) _:udef_q($\boxed{x_2}$, $\boxed{h_5}$, _) }, { $\boxed{h_5} =_q \boxed{h_1}$ } ⟩
→ ⟨$\boxed{h_0}$,{$\boxed{h_3}$:_($\boxed{x_4}$) }, {} ⟩

n_n_lexicalization_mtr ∧ { tur_n, gåer_n } → { hiker_n }

Again, a high-level transfer type accounts for a general translational equivalence, viz. the translation of a SL nominal compound into a TL lexical noun. The standard MRS analysis of compounds is by means of a two-place 'unspec' relation (as in most cases the grammar has little to say about the specific relation between the two parts), acting much like an intersective modifier on the head noun; the non-head, in turn, is bound by an underspecified indefinite quantifier, since our use of MRSs assumes that all instance variables need to be uniquely bound by a quantifier. The MTR type consumes four input EPs plus the handle constraint associated with the covert quantifier and outputs a single EP, the relation of the TL lexicalization, while preserving the handle and instance variable of the compound head on its output. A sample instantiation of this type is the translation of Norwegian *turgåer* ('hike-walker') into English *hiker* (note that, had the grammars decided to lexically decompose agentive nominalizations, the pattern would still hold true, modulo the additional EP for the underlying verbal predicate). Making this rule obligatory ensures that word-by-word translations of *turgåer* are blocked.

A final MTR example demonstrates the use of predicate typing in transfer, viz. the translation of temporal prepositions, where the selection of a specific preposition appears to be largely determined by language-internal constraints.

arg12_mtr ∧ { temp_abstr($\boxed{x_0}$) } : { på_p(_, $\boxed{x_0}$) } → { temp_loc }
arg12_mtr ∧ { på_p } $\overset{?}{\rightarrow}$ { on_p }
arg12_mtr ∧ { på_p } → { at_p }

---

[3] To avoid the costly full semantic comparison of each newly derived MRS against the set of earlier intermediate results, MRSs are organized in a chart of sorts, using a bit-vector coded representation of transfer derivations for indexing and retrieval of candidate equivalent MRSs.

Aggregate	**total items** $\sharp$	**word string** ϕ	**distinct trees** ϕ	**overall coverage** %	**time (s)** ϕ
$30 \leq \mathit{words} < 35$	2	32.50	56.00	100.0	47.26
$25 \leq \mathit{words} < 30$	7	26.71	36.67	85.7	21.09
$20 \leq \mathit{words} < 25$	26	21.58	153.12	92.3	15.75
$15 \leq \mathit{words} < 20$	68	16.88	62.06	100.0	5.44
$10 \leq \mathit{words} < 15$	117	11.77	6.78	100.0	1.12
$5 \leq \mathit{words} < 10$	90	7.54	3.08	100.0	0.39
$0 \leq \mathit{words} < 5$	6	4.00	1.17	100.0	0.20
Total	**317**	**12.86**	**29.73**	**98.7**	**3.63**

(generated by [incr tsdb()] at 16-mar-2004 (12:51 h))

Table 1: Central measures of generator performance (for English translations of the current development corpus) 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.

By virtue of its context condition, the first rule will transfer Norwegian *på* into the abstract predicate ‘temp_loc’ (see Figure 4 above), if and only if its internal argument is of a temporal sort. Letting the generation grammar determine the choice of TL preposition, subsequently, ensures that Norwegian *på ettermiddagen* (‘on afternoon$_{\text{def}}$’) is correctly translated as English *in the afternoon*.

Transfer rules are applied until the rewrite process reaches a fix-point, which can be guaranteed as long as the transfer grammar avoids direct or indirect rewrite cycles. In order to obtain a notion of completeness of output MRSs, EPs are ‘color-coded’ during transfer: for the time being, we employ two colors to indicate SL and TL status, respectively (with a small number of chameleon-type, interlingual predicates like ‘message’ EPs representing illocutionary acts; see Ginzburg & Sag, 2000). Individual MTRs (or often sets of rules) indicate whether they act color-preserving (i.e. SL to SL or TL to TL) or color-changing (SL to TL), thus effectively creating independent namespaces for predicates and enabling individual transfer phases to optionally filter outputs that have a non-zero intersection with the input predicate set. The final transfer result(s) can be tested for compatibility against the published interface to the generation grammar based on the target language SEM-I, which goes some way towards ensuring generator success.

6 English Generation using HPSG

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). After addition of domain-specific vocabulary and a small amount of fine-tuning, the ERG provides adequate analyses for around 95 per cent of the 330-sentence corpus of translations for the original Norwegian development corpus. Yet, out-of-the-box LKB generator performance was not satisfactory at an average processing time of above 40 seconds per input MRS and frequent generator time-outs.

In-depth study of generator performance revealed two major computational sinks, viz. indexing on subsets of semantic relations and exponential combinatorics originating in local ambiguity. While the first was easily addressed by superimposing a bit-vector based indexing scheme, the second led to a revision of the basic chart generation algorithm, blending subsumption-based packing in the spirit of Oepen & Carroll (2002) with the original two-phase treatment of modifier attachment. We plan to publish the new algorithm and further evaluation results separately, but Table 1 indicates that generation complexity from well-formed MRSs could be reduced to approximate polynomial growth in input complexity and time-outs eliminated completely.

At an average of 30 English realizations per input MRS (see Table 1), it is vital to rank the generator output. As a first shot at deriving a likelihood measure for realizations, strings are assigned perplexity scores with respect to n-gram language models. Using the CMU SLM toolkit with Witten-Bell discounting and a vocabulary of 65,000 words, models have been trained for various values of n on the British National Corpus. Our best performer to date, a 5-gram model with back-off, achieves an exact match accuracy of around 55 per cent (against a 30 per cent random-choice baseline) and quite generally seems to prefer idiomatic-sounding realizations over less acceptable alternatives. Further experimentation with structural features and different learners is now underway.

7 Satellite Activities: Evaluation & Resources

Complementing work on the core MT prototype, LOGON pursues more basic, PhD-level research (on disambiguation techniques, soft constraints, WSD, and the syntax – semantics interface) as well as resource creation—adaptation of a large computational lexicon and associated tools and the production of a parallel domain corpus—and evaluation activities. For component-level evaluation we have revised the glass-box *competence and performance profiling methodology* (Oepen & Carroll, 2002) for both transfer and generation (see examples in § 4 and § 6), but also foresee a round of end-to-end, black-box evaluation to assess the utility of currently fashionable, n-gram based similarity metrics.

8 Preliminary Conclusion — Outlook

The idiom *Som å kapp-ete med trollet* to many Norwegians suggests an intricate task that demands creative thinking. The future of MT has been (mis-)diagnosed 'just around the corner' since the beginning of time. In absolute numbers LOGON only commands five per cent of the resources available to Verb*mobil*; there is no basis to expect a break-through in fully-automated MT in the foreseeable future, but yet we see progress along the way, specifically in the sustained development of large-scale, general purpose language technology and a growing understanding of the role of practical computational semantics. In only nine months of development, the LOGON consortium assembled a functional end-to-end prototype, albeit of limited coverage, a novel integration of LFG analysis and HPSG generation, and an innovative approach to ambiguity-preserving transfer—all in a scalable architecture. We expect to broaden the scope of our prototype continually, specifically in terms of transfer coverage, and in parallel plan to pursue a few in-depth feasibility studies, for example on the use of more 'geometric' semantic accounts of aspectual and temporal relations or modal operators.

Extending ambiguity management throughout all translation phases, we expect to devise stochastic models to reduce the number of active hypotheses upon completion of each phase—i.e. parse selection in the spirit of Riezler, Prescher, Kuhn, & Johnson (2000), ordering transfer outputs on the basis of log-linear models trained on an English MRS bank, and more refined probabilistic models to rank generation results—as well as for end-to-end selection of resulting translations (e.g. using a crude SMT model *on top* of the symbolic backbone). At the same time, we are actively pursuing the model of 'chunk translation' sketched in § 4, combined with optional inclusion of some syntactic information on surface ordering, to mitigate the inherent brittleness to out-of-scope input, regarding both the analysis and generation grammars. To eventually broaden transfer coverage, we expect to leverage a broad-coverage bilingual dictionary for basic translational equivalences at the lexical level. Snipp, snapp, snute—så er eventyret ute.

References

Alshawi, H., & Crouch, R. (1992). Monotonic semantic interpretation. In *Proceedings of the 30th Meeting of the Association for Computational Linguistics* (pp. 32 – 39). Newark, NJ.

Bond, F., Ogura, K., & Ikehara, S. (1996). Classifiers in Japanese-to-English machine translation. In *Proceedings of the 16th International Conference on Computational Linguistics* (pp. 125 – 130). Copenhagen, Denmakr.

Bos, J. (1995). Predicate logic unplugged. In *Proceedings of the tenth Amsterdam colloquium* (pp. 133 – 142). Amsterdam, The Netherlands.

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.

Copestake, A. (2002). *Implementing typed feature structure grammars.* Stanford, CA: CSLI Publications.

Copestake, A., Flickinger, D., Malouf, R., Riehemann, S., & Sag, I. (1995). Translation using minimal recursion semantics. In *Proceedings of the Sixth International Conference on Theoretical and Methodological Issues in Machine Translation.* Leuven, Belgium.

Copestake, A., Flickinger, D., Sag, I. A., & Pollard, C. (2003). *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.

Dyvik, H. (2003). ParGram. Developing parallel grammars. *ELSNews*, *12*(2), 12 – 14.

Egg, M., Koller, A., & Niehren, J. (2001). The constraint language for lambda structures. *Journal of Logic, Language, and Information*, *10*, 457 – 485.

Flickinger, D. (2000). On building a more efficient grammar by exploiting types. *Natural Language Engineering*, *6 (1) (Special Issue on Efficient Processing with HPSG)*, 15 – 28.

Ginzburg, J., & Sag, I. A. (2000). *Interrogative investigations. The form, meaning, and use of English interrogatives.* Stanford, CA: CSLI Publications.

Oepen, S., & Carroll, J. (2002). Efficient parsing for unification-based grammars. In S. Oepen, D. Flickinger, J. Tsujii, & H. Uszkoreit (Eds.), *Collaborative language engineering. A case study in efficient grammar-based processing* (pp. 195 – 225). Stanford, CA: CSLI Publications.

Reyle, U. (1993). Dealing with ambiguity by underspecification. Construction, representation, and deduction. *Journal of Semantics*, *10*, 123 – 179.

Riezler, S., Prescher, D., Kuhn, J., & Johnson, M. (2000). Lexicalized stochastic modeling of constraint-based grammars using log-linear measures and EM training. In *Proceedings of the 38th Meeting of the Association for Computational Linguistics* (pp. 480 – 487). Hong Kong, China.

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