

Interactive MT As Support For Non-Native Language Authoring

Svetlana Sheremetyeva and Sergei Nirenburg

Computing Research Laboratory
New Mexico State University

Abstract

The paper describes an approach to developing an interactive MT system for translating technical texts on the example of translating patent claims between Russian and English. The approach conforms to the human-aided machine translation paradigm. The system is meant for a source language (SL) speaker who does not know the target language (TL). It consists of i) an analysis module which includes a submodule of interactive syntactic analysis of SL text and a submodule of fully automated morphological analysis, ii) an automatic module for transferring the lexical and partially syntactic content of SL text into a similar content of the TL text and iii) a fully automated TL text generation module which relies on knowledge about the legal format of TL patent claims. An interactive analysis module guides the user through a sequence of SL analysis procedures, as a result of which the system produces a set of internal knowledge structures which serve as input to the TL text generation. Both analysis and generation rely heavily on the analysis of the sublanguage of patent claims. The model has been developed for English and Russian as both SLs and TLs but is readily extensible to other languages.

1 Introduction

Translating patents is an important task for international trade and industry. In a patent text (or "disclosure," using official terminology), the crucial part is the patent claim which is the actual subject of legal protection. Analysis synthesis and translation of claims are time-consuming tasks even for experts. The initial functionality of our patent specialist's workstation (Sheremetyeva and Nirenburg 1996b) was centered on claim text composition. The project we describe here extends that functionality to translation of patent claims.

General purpose MT systems have the advantage of being potentially reusable; this reusability is not, however, guaranteed. A study by Bourbeau and Kittredge (1988) demonstrated that no existing NLP system could process patent texts adequately. It is generally recognized, however, (see, e.g., Cowie and Lehnert 1996) that an MT system providing adequate performance even for a single type of text should be considered useful. If an MT system

uses a restricted sublanguage—and, thus, can operate with smaller-scale static knowledge sources—the scope of acquisition and development effort will decrease correspondingly. Indeed, practically all MT systems for special domains are usually (see, e.g., Kukich 1983; Kittredge *et al.*, 1986) built to conform to the constraints of a sublanguage.

Massive attempts have been made in the past ten years or so to make MT systems fully automatic (e.g., P. Brown *et al.*, 1988). In practice, the state of the art in NLP suggests a mixture of automatic and manual methods for any realistic comprehensive application (e.g., Paris *et al.*, 1995; Nirenburg *et al.*, 1996). Several modes of human-computer cooperation have been used in practice over the years. Our approach conforms to the human-aided machine translation (HAMT) paradigm (e.g., Kay 1973, see also Hutchins and Somers 1992 for a definition). We would also like to stress an additional important parameter of human-computer interaction in HAMT: *initiative*. Human-computer interaction in HAMT system could be initiated either by the system or by the human (sometimes both modes are present in a single application). In our model, the initiative is predominantly, though not exclusively, with the system.

In this paper we describe a method for developing an interactive domain-tuned Russian↔English HAMT environment for patent claims on apparatuses which is a part of a workstation for a multilingual processing of patent texts¹. (In the description, we will concentrate on the Russian-English direction, though the method is identical for both directions.) This module is a tool for a user (an inventor or patent officer) who is a Russian speaker and does not necessarily know English.

2 System Overview

The system takes a Russian claim text as input and outputs its English translation in the format which meets all

1. Among other functionalities of the workstation will be information retrieval, information extraction, translation between the languages, patent disclosure generation. A module for generating patent claims in English is described in Sheremetyeva and Nirenburg (1996a) and Sheremetyeva and Nirenburg (1996b).

US legal requirements for patent claims. Examples of rather simple Russian and US parallel claim texts describing apparatuses (which can be more than a page long) are given in Figures 1 and 2.

*Greifer, soderzhaschii traversu, dve pary cheljustej, sharnirno soedinennye mezhdu soboj obschei osju i posredstvom tjagi — s traversoj, i raspolozhennuju na traverse polispastuju sistemu, kanat kotoroj svjazan s lebedkoj i **propuschen** cherez bloki, **ustanovlennye** na obschei osi soedinenija cheljustej, o t I i ch a ju sch e i s ja tem, chto lebedka **smontirovana** na odnoj cheljusti vblizi obschei osi, prichem **vzaimodeistvujuschij** co vtoroj cheljustju podpruzhinennyj T-obraznyj ryhag sharnirno **zakreplen** na soedinennoj so vtoroj cheljustju tjage.*

FIGURE 1. An example of a Russian patent claim text. Predicative words which are heads of individual phrases describing essential features of the invention are bold faced.

*A clamshell **comprising** a traverse, two pairs of jaws **coupled** therebetween pivot ally by common axis and **connected** to the traverse by means of pulls, a pulley block system **disposed** on the traverse, the rope of said traverse **being associated** with a winch and **being passed** across the blocks which **are mounted** on the common axis of jaws connection *characterized in that* the winch **is mounted** on one jaw in proximity to the common axis, a spring-actuated T-shaped lever **cooperating** with the second jaw and pivotally **being held** on the pull **connected** to the second jaw.*

FIGURE 2. An English translation of the Russian patent claim text presented in Figure 1. This translation meets all legal requirements to a claim text. Predicative words which are heads of individual phrases describing essential features of the invention are bold faced.

Patent claims must be formulated as specified by the German Patent Office and commonly accepted in the U.S., Russia and other countries. The claim must describe essential features of the invention in the obligatory form of a single extended nominal sentence with a well-specified conceptual, syntactic and stylistic/rhetorical structure which frequently includes long and telescopic embedded predicate phrases. The generic features of the invention must be described first, followed by the "difference" (novelty) features. The generic and difference parts of a patent claim are connected by the fixed expression *characterized in that* (*otlichajuschisja tem, chto* in Russian). So as best to protect the rights of the inventor, it is desirable to use lexical units whose meanings are as broad as possible. These requirements apply to the description of all types of inventions recognized by the U.S. and Russian Patent Laws (devices, substances, methods, living organisms, etc.), though the lexicons and some morphological and syntactic features are specific for every type of a claim. The HAMT

system for patent claims consists of:

- an analysis module which includes i) a submodule of *interactive syntactic analysis* of SL text (decomposition of a syntactically complex nominal sentence into a set of simple structures equivalent to predicate phrases describing individual features of an invention) and ii) a submodule of *fully automated morphological analysis* of the word occurrences in these simple structures².
- an automatic module for transferring the lexical and partially syntactic content of SL text into a similar content of the TL text;
- a fully automated TL text generation module which relies on knowledge about the legal format of TL patent claims.

3 The Background Knowledge

For successful translation of patent texts two distinct types of expert knowledge are necessary: knowledge about the sublanguage of patents as legal documents and knowledge about the technical field of the invention. The legal knowledge essentially makes itself manifest in the constraints on and preferences concerning claim syntax (though it also affects lexical elements). The technical knowledge is mainly conveyed by domain-tuned terminology. Both kinds of knowledge are encoded in the system lexicon.

One of the characteristic features of our system is that it reuses the domain-tuned knowledge and knowledge representation language (Sheremetyeva 1999) and the automated generation module of an implemented interactive computer system for authoring patent claims for an English speaker. (Sheremetyeva and Nirenburg 1996a; Sheremetyeva and Nirenburg 1996b) as well as the Russian morphological analyzer developed and implemented for the Corelli MT project (Sheremetyeva and Nirenburg 1997). This knowledge was augmented by bi-lingual dictionaries and transfer rules mapping shallow knowledge representations between SL and TL (Russian and English in our case). The background knowledge for the system includes:

- a shallow bilingual (Russian/English) lexicon of nominal terminology which is, in fact, what Heid and McNaught (1991: 35) call a reusable resource of the first kind: a resource which was once build for some other purpose, already exists on-line and can be simply fed into the system. This dictionary includes lexical units simply listed with their class membership which is an MT-oriented semantic classification of groups of words and phrases with similar syntactic properties, and

2. The early application of syntactic analysis allows the morphological analyzer to avoid overgeneration and produce unambiguous results.

- deep (information-rich) bilingual lexicons of predicates (heads of predicative phrases describing essential features of an invention) used in US and Russian patent claims; these lexicons have been specifically constructed for the application and are meant for a multi-functional use in other modules of the patent workstation. These lexicons are the main part of the system knowledge and can be called a reusable resource of the second kind in the sense of Heid and McNaught (1991: 35).

The system contains predicate lexicons for both SL and TL. This is the essential part of the system knowledge which covers both the lexical and, crucially for our system, the syntactic knowledge. Our approach to syntax is, thus, fully lexicalist (cf. Ooi 1998: 6). The set of parameters (or fields) for predicate specification in our knowledge base is strictly determined by the needs of application and draws heavily on the sublanguage corpus analysis in both English and Russian.

The entry in a predicate lexicon is organized as follows:

dictionary::= {entry}+

entry::= major-form other-forms semantics freq case-frame patterns translation

major-form the most frequent morphological form of the predicate in which it occurs in patent claims; used in text generation for choosing the morphological form of a predicate; helps to simplify morphological generation of word forms in the output text;

other-forms morphological forms of the verb in which it occurs in patent text; used for the same purpose as the knowledge in the previous field as well as to search predicates in the claim text during interactive analysis;

semantics the verb's semantic class (the values are taken from a predefined set. e.g., *meronymic*, *spatial*, etc.). This information is used at the generation stage to determine the order in which the predicates should appear in the text;

freq the predicate's frequency rank in the list of the predicates from the sublanguage corpus which belong to one of the above semantic classes; used to estimate the breadth of predicate meaning, an important preference feature in patent claim composition;

case-frame the set of the verb's case roles, with their ranks, that is, their relative importance for the given predicate, as estimated by the frequency of their co-occurrence with this predicate in a corpus; used at all stages of translation;

patterns a list of alternative linearization patterns for the verb's case frame, in the order of decreasing frequency of occurrence of the verb with a particular subset of case roles: for example, the following phrase from an actual claim: (1: *the splice holder*) *: *is mounted* (2: *on the cover part*) (4: *to form a rotatable splice holder*) (where 1, 2 and 4 are case role ranks and "*" shows the position of the predicate) will match the linearization pattern (1 * 2 4);

translation crosslinguistic equivalents; if no transfer conditions are specified the equivalent specified is the default translation.

A sample entry in the English predicate lexicon is shown below.

mounted	"mounted" F
(major-form	((("is mounted"))("are mounted"))
more-forms	("being mounted"))
sem-class	location
freq	1
case-frame	((1 subject)(2 place)(3 manner)
	(4 purpose)(5 means))
patterns	((1 * 2)(1 3 * 2)(1 * 2 4)(1 * 2 3)(1 * 3)
	(1 * 4)(1 * 2 5)(1 3 * 2 4)(1 * 3 * 4)
	(3 * 1)(1 3 * 2 3)
translation	("ustanovlennyj")

4 Interactive Analysis of a Claim

The goal of the interactive stage of claim analysis is to elicit from the user (who is a speaker of Russian and may not know English) conceptual knowledge about the structure of the invention and linguistic knowledge about the syntactic structure of the Russian claim text to be translated so as to make further stages of translation procedure completely automatic. Drawing on the knowledge it has about the kinds of information typically presented in patents and using common graphical user interface tools (such as dialogue boxes, menus, templates, slide bars etc.) the system guides the user through the paces of "understanding" the structure of an invention and the text by decomposing a complex input text into predicate phrases describing individual features of the invention "disguised" in the complex telescopic claim structure. The interactive analysis scenario is described by the following algorithm:

```
begin
    elicit-type;
    elicit-predicate-phrases:
        elicit-predicate;
        elicit-case-role-fillers;
        mark-co-references;
end
```

During the **elicit-type** procedure the system is automatically tuned to a particular type of a claim. This is based on a very simple heuristics that the title of the claim often contains a genus term owing to the requirement of breadth of reference which enhances protection against patent infringement. The genus terms for apparatuses include such terms as *apparatus*, *construction*, *assembly*, *device*, *means*, *machine*, *unit*, etc., but such terms as *method*, *process*, *organism*, etc.. which are genus for other types of claims are not allowed. If no genus term is found in the title the system treats the invention as apparatus based on the heuristic that it is the most numerous type of invention.

The set up for this first elicitation procedure involves displaying a title and a text of a patent claim in SL (which can be either scanned or typed in by the user) and a menu featuring all types of inventions recognized by the U.S. Patent Office (devices, substances, methods, living organisms, etc.) with the highlighted type of an invention described in the claim. The user can either accept the type or click on a different type value in the menu. Once the selection is made, the system selects the appropriate stages of the syntax elicitation (and generation) process and the corresponding lexicons to support the process. The procedure mark-co-references applies to patents of all types.

then fills the slots of this template with text elements by highlighting appropriate words or phrases in the claim text and pasting them into the template. The phrases filling the case roles are treated as constituents, yielding output of the required syntactic analysis. (Knowing the boundaries of phrases helps fight overgeneration in morphological analysis). Once the "Submit" button is pressed after the template is filled, the system produces the internal representation of a predicate phrase. In Figure 3, we illustrate the interactive syntactic analysis on an English example, for readability.

Note that this method of syntax elicitation has the advantage of dealing with ellipsis in a very simple way. The

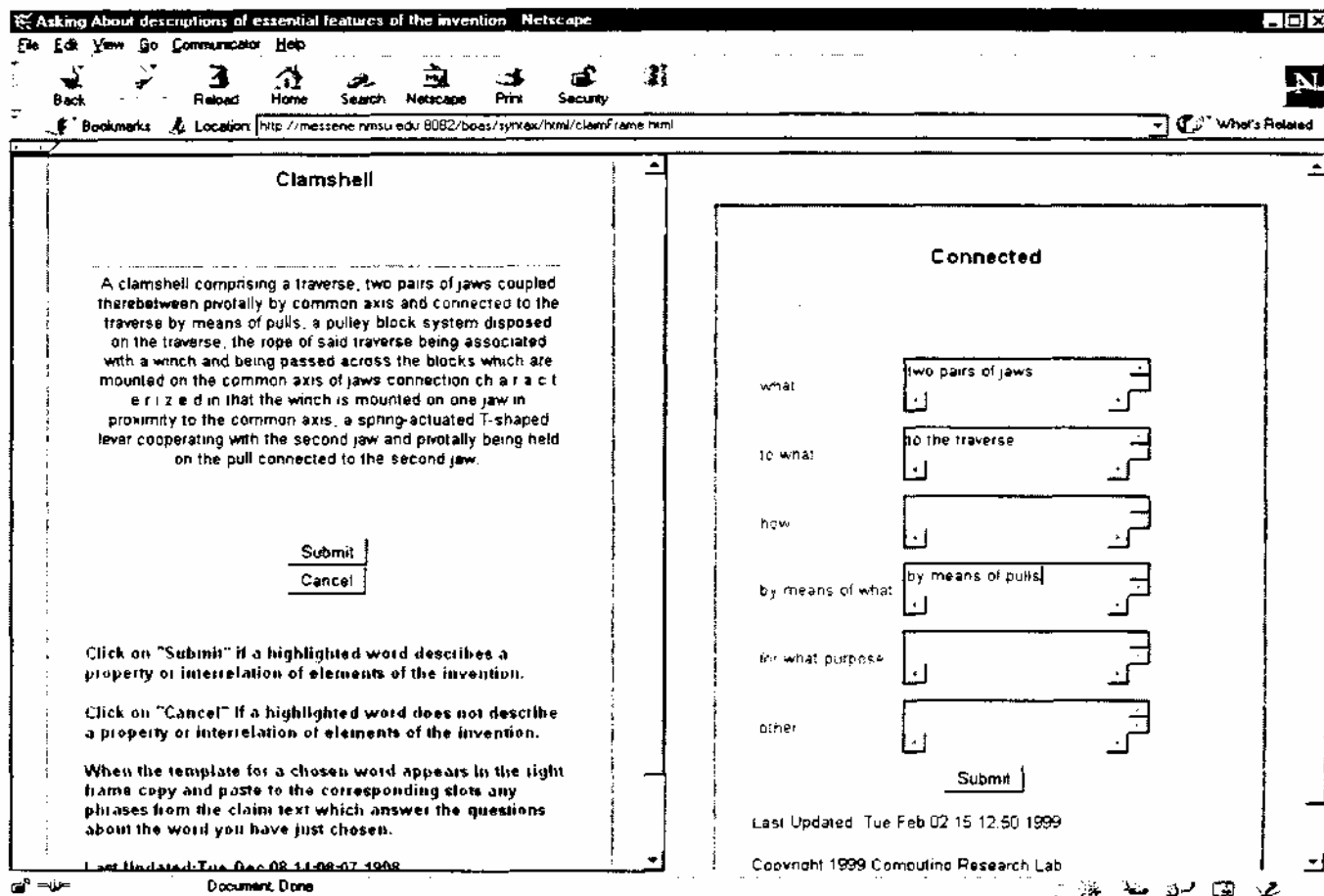


FIGURE 3. Filling case roles in a predicate template.

The next step is to elicit individual predicate phrases from the telescopic structure of an input claim. This is done with the help of the **elicit-predicate-phrases** procedure.

Along with the questions and the instructions the system highlights³ all the predicates in the text, one at a time. The user has a choice of rejecting a predicate suggested by the system by clicking on the "Cancel" button (if this word is not the head of a predicate phrase describing a feature of the invention) or accepting it, by clicking on the "Submit" button. Once a predicate is selected, a template based on its case roles is displayed in a separate frame. The user

system is treating every dash in a claim text as a predicate. When the user accepts a dash as a predicate, a copy of the

3. This is feasible, as the inventory of predicates that can appear in a claim is rather small. An analysis of more than 1,000 US patent claims and a similar number of Russian claims on apparatuses showed that in the US corpus, 98% of surface predicates were covered by only 531 lexemes. The Russian sublanguage is even more restricted—98% of all predicate word forms are covered by only 65 predicate lexemes.

template corresponding to the previous predicate in the text is displayed. For example, in the Russian claim text in Figure 1 the dash "—" substitutes the predicate "soedidinennye" (connected) and the content of this invention feature is represented as the following template:

(p-generic 3 "soedineny" (connected)
 (1 subject "dve pary cheljustej" (two pairs of jaws)
 (2 object "s traversoj" (to the traverse)
 (4 means "posredstvom tjag" (by means of pulls)))

FIGURE 4. Internal filled predicate template.

If a predicate is found in the text before the fixed expression "characterized in that" it is marked as "generic" otherwise as "novelty".

This procedure halts when all the properties of and relations among invention pans are elicited. To facilitate target claim text generation, at this stage all co-references to objects mentioned in the templates are established. In the main authoring window, the system highlights all the sets of co-reference candidates, set by set. The user is asked to remove any elements from the candidate sets that are not co-referential with the rest by clicking on the "Cancel" button. Once this procedure finishes, all the information necessary for automated claim translation has been elicited.

The output of this stage of the patent expert's workstation operation consists of a knowledge structure which is a set of filled predicate templates representing the technical content of a claim text (see Figure 5).

```
text          ::= {template} {template}*
template     ::= (predicate-class predicate( {case-role}
{case-role}*))
case-role    ::= (rank (value))
```

FIGURE 5. Internal claim text representation.

Predicate-class is the label of a synonym set of predicate-type words, **predicate** is a string corresponding to a predicate from the system lexicon, case roles are ranked based on their frequency of co-occurrence with each predicate in the training corpus and **value** is the string which fills a case role. Note that the order of the templates is not relevant. The predicates have "generic" or "novelty" markers.

The last stage of the analysis module is part-of-speech tagging of case-role fillers which is done using an off-the-shelf Russian morphological analyzer (Sheremetyeva and Nirenburg 1997).

5 From Interactive Knowledge Elicitation to Claim Text Generation

5.1 Cross-linguistic Equivalence

Patent sublanguages differ from standard natural languages both in their sense inventory and typical predicate/

argument structures. This affects the choice of cross-linguistic equivalents, and therefore the procedure of defining these equivalents has been corpus-based. A parallel analysis of US claims and their Russian translations from "The Official Gazette" published in Russia was carried out and translinguistic equivalents were extracted for every English/Russian predicate. For example, the Russian predicate *ustanovlennyyj* could be used to translate the following English predicates: *mounted, supported, disposed, arranged, positioned, fixed, carried, maintained, received, placed, located, fitted, opposed*. This raises the question of whether these equivalents can be used in translation indiscriminately and if they cannot, what criteria govern the choice of a translation.

The choice of translinguistic equivalents based only on sense proximity often leads to restructuring problems which in NLP can be even more difficult than homonymy resolution. To bypass this problem we argue that the equivalence criteria should also take into account proximity of the structural properties of lexemes. In the system described, the choice of English/Russian predicate equivalents out of the set of synonyms is based on the similarity among the values of following parameters: frequency, case-role structure, case-role rank, case-role morphological representation and linearization patterns.

Different sets of values of the above parameters for one and the same predicate can lead to choosing different equivalents. For example, the Russian *ustanovlennyyj* translates into *mounted* if it appears with the case roles of ranks 1 and 2 and as *arranged* if it appears with the case roles of ranks 1 and 3. If we have a three case-role linearization pattern, the co-occurrence of case roles ranked 1, 2 and 3 or ranked 1, 2 and 4 yields the English equivalents "disposed" or "positioned," respectively. The ranks of case roles and linearization patterns are not always the same across languages:

- 1: blok instrumentov *:ustanovlen 2:na osnovanii = 1 : tooling means *:is **mounted** 2:on said base means;
- 1: pervyj nozh, *:ustanovlennyyj 3:poperek prodolxnoj osi slitka = 1:first blade *:arranged 3:transverse to the longitudinal axis of the ingot;
- 1:rychagi *:ustanovleny 2:pod ramoj 3:vertical'no = 1:said arms **are** 3:vertically **disposed** 2:below said frame;
- 1: zatvor 4: s pomoshchju shaiby *:ustanivlen 2:na rastjagivaemyh elementah = 1:said bolt *:is **positioned** 2:on said tension members 5:by the spacer;

The Russian/English predicate equivalents thus selected convey the same sense and eliminate (or at least reduce to a minimum) the need for phrase restructuring after transfer. The transfer rules of the system are formulated in terms of case-role linearization patterns. The TL equivalents for case-role fillers are found in the regular domain tuned terminology dictionaries with the usual MT routine procedures.

5.2 Transfer

The input for this stage is a set of Russian predicate templates with part-of-speech tagged case-role fillers. The output is a set of equivalent English predicate templates with part-of-speech tagged fillers (see examples in Figure 6).

```
[(p-generic 3 "soedineny"
  (1 (Num1 N1 N2) "dve pary cheljustej")
  (2 (Adv1) "mezhdu soboj")
  (3 (Adv2) "sharnirno")
  (4 (Prep2 N4) "posredstvom tjag"))]
(p-generic 3 "coupled"
  (1 (Num1 N1 Prep3 N2) "two pairs of jaws")
  (2 (Adv1) "therebetween")
  (3 (Adv2) "pivotally")
  (4 (Prep2 N4) "by means of pulls")))]
[(p-genenc 3 "soedineny"
  (1 (Num1 N1 N2) "dve pary cheljustej")
  (2 (Prep1 N3) "s traversoj")
  (4 (Prep2 N4) "posredstvom tjag"))]
(p-generic 3 "connected"
  (1 (Num1 N1 Prep3 N2) "two pairs of jaws")
  (2 (Prep1 N3) "to the traverse")
  (4 (Prep2 N4) "by means of pulls")))]
```

FIGURE 6. Examples of template transfer from Russian-oriented into English-oriented.

The transfer algorithm works as follows. First, the SL case-role fillers are substituted with their equivalents in TL, with the help of available glossaries and dictionaries. Next, SL predicates are substituted with their TL equivalents according to the equivalence criteria described in the previous section. Note that the Russian predicate "soedineny" is transferred into more than one English predicate, due to variations in case-role structure.

Any morphological features of predicates, the constituent order within predicate phrases and the order of predicate structures in the claim text are not transferred. These parameters of the output text are determined during the generation stage based on the information in the English predicate lexicon. In our example (see Figures 1 and 2), the same order of predicates in the source and target texts is due to the similarity of rhetorical and stylistic requirements of the patent claim sublanguages of SL and TL.

6 Automatic Generation

The claim text generation stage takes an English-oriented text representation (Figure 5) as input and produces a text of the claim in a legal format, as illustrated in Figure 2.

As was already mentioned above, our MT module fully reuses an automatic generator described in detail in Sheremetyeva and Nirenburg (1996b). Superficially, the generator architecture conforms to the standard emerged in natural language generation (NLG) (as expressed, for instance in Reiter 1994), in that it includes the stages of content specification, text planning and surface generation (realiza-

tion). The results of the transfer stage (two lists, generic and novel, of filled TL predicate templates which specify the content of the claim) are submitted to an automatic text planner which outputs an hierarchical structure of templates. The planning stage is guided both by constraints on the patent claim sublanguage and the general constraints on style. The former determines the global ordering of the claim text while the latter deals with local text coherence. This process resembles revision-oriented generation (Meter 1991, Robin 1994, Gabriel 1988, Inui *et al.*, 1992). The realization stage of the generator linearizes the hierarchy of TL predicate templates of each group and takes care of the ellipsis, conjoined structures, punctuation and morphological forms. The two completely ready parts of the claim text are bound by the intermediate expression *characterized in that*, the generic and novelty parts being put correspondingly before and after this expression.

7 Status and Future Work

The system is in the late stages of implementation as of April 1999. The static knowledge sources—the dictionaries for both languages, including transfer-related knowledge—have been compiled for the sublanguage of patents about apparatuses. The morphological analysis of Russian is operational and well tested. The English generator is also operational. The interactive syntactic analysis has been implemented using the technology developed in the Boas project at NMSU CRL (e.g., Nirenburg and Raskin 1998). By the time of the conference, the system will be integrated and tested.

We intend to a) extend the system into multilingual generation and machine translation. Another direction of work is developing the interactive authoring support with human-computer interaction in a variety of languages (this could be called "software localization"); b) develop a patent search facility on the basis of the patent disclosure sublanguage and the information retrieval and extraction infrastructure developed in the TIPSTER project (Grishman, 1995).

References

- Bourbeau L. and R. Kittredge. 1988. *Project d'automatisation aux brevets et inventions: traduction assistée par ordinateur (rapport final)*. Ministère de la Consommation et des Corporations Canada. Direction des Systèmes Automatisés. Bureau de la propriété intellectuelle.
- Brown P., J. Cocke, S. Della Pietra, V. Della Pietra, F. Jelinek, R. Mercer and P. Roossin. 1988. *A Statistical Approach to Language Translation*. Proceedings of COLING-88. Budapest, pp. 71-74.
- Cowie J. and Lehnert W. 1996. *Information Extraction*. In *Communications of the ACM*. Vol. 39, No.1, January, pp. 80-91.

- Gabriel R. 1988. *Deliberate writing*. In D.D. McDonald and Bolc L., editors. **Natural Language Generation Systems**. Springer-Verlag, New York, NY.
- Grishman R. 1995. *Tipster Phase II Architecture Design Document, version 1.52*. TIPSTER Architecture Working Group.
- Heid U. and McNaught J. 1991. *EUROTRA-7 Study: Feasibility and Project Definition Study on the Reusability of Lexical and Terminological Resources in Computerized Applications, Final Report*. Luxemburg: Commission of the European Communities.
- Hutchins W.J., and H.L. Somers. 1992. *An Introduction to Machine Translation*. Cambridge: Cambridge University Press.
- Inui, K., Tokunaga, T., and Tanaka, H. 1992. *Text revision: a model and its implementation*. In Dale, R., Hovy, E., Roesner, D., and O. Stock, editors. **Aspects of Automated Natural Language Generation**, pp. 215-230. Springer-Verlag.
- Kay M. 1973. *The MIND System*. Natural Language Processing. New York: Algorithmic Press.
- Kittredge, R., Polguère, A., and Goldberg. E. 1986. *Synthesizing weather forecasts from formatted data*. In Proceedings of the 11th International Conference on Computational Linguistics, pp. 563-565. COLING.
- Kukich K. 1983. *Knowledge-Based report generation: a knowledge engineering approach to natural language report generation*. Ph.D. thesis. University of Pittsburgh.
- Meteer M.W. 1991. The implications of revisions for natural language generation. In Paros, C., Swartout, W., and Mann, C., editors. *Natural Language Generation in Artificial Intelligence and Computational Linguistics*. Kluwer Academic Publishers, Boston.
- Nirenburg S., Beale S., Helmreich S., Mahesh K., Viegas E., Zajac R. 1996. Two Principles and six techniques for Rapid MT Development. *Proceedings of the Second Conference of the AMTA*. Montreal.
- Nirenburg S. and V. Raskin. 1998. Universal Grammar and Lexis for Quick Ramp-Up of MT Systems. *Proceedings of COLING-ACL '98*. Montreal.
- Ooi B.Y. 1998. *Computer Corpus Lexicography*. Edinburgh University Press.
- Paris C., Vander Linden K., Fisher M., Hartley A., Pember-ton L., Power R., and Scott D.A. 1995. Support Tool for Writing Multilingual Instructions. *Proceedings of IJCAJ-95 V.2* Montreal.
- Robin J. 1994. *Revision-Based Generation of Natural Language Summaries. Providing Historical Background. Corpus-Based Analysis, Design, Implementation and Evaluation*. Technical Report CUCS-034-94.
- Sheremetyeva S. and S. Nirenburg. 1996a. *Interactive Knowledge Elicitation in a Patent Expert's Workstation*. IEEE Computer. Vol.7.
- Sheremetyeva S. and S. Nirenburg. 1996b. *Generating Patent Claims*. Proceedings of the 8th International Workshop on Natural Language Generation. Herstmonceux, Sussex, UK.
- Sheremetyeva S. and S. Nirenburg. 1997. *Minimizing Acquisition and Development Effort in Computational Morphology*. Proceedings of CLIN-97. Nijmegen. December.
- Sheremetyeva S. *A Flexible Approach To Multi-Lingual Knowledge Acquisition For NLG*. 1999. Accepted for presentation at the 7th European Workshop on Natural Language Generation. Toulouse. (France) May 13-15.