

# Interactive Resolution of Intrinsic and Translational Ambiguity in a Machine Translation System\*

Igor M. Boguslavsky<sup>1,2</sup>, Leonid L. Iomdin<sup>1</sup>,  
Alexander V. Lazursky<sup>1</sup>, Leonid G. Mityushin<sup>1</sup>, Victor G. Sizov<sup>1</sup>,  
Leonid G. Kreydlin<sup>1</sup>, and Alexander S. Berdichevsky<sup>3</sup>

<sup>1</sup> Institute for Information Transmission Problems, Russian Academy of Sciences  
Bolshoj Karetnyj pereulok, GSP-4,  
Moscow 129994, Russia  
{iomdin, lazur, mit, sizov, lenya}@iitp.ru  
<http://proling.iitp.ru>

<sup>2</sup> Madrid Polytechnic University, Faculty of Informatics  
Campus de Montegancedo S/N  
28660 Boadilla del Monte, Madrid, Spain  
igor@opera.dia.fi.upm.es

<sup>3</sup> Moscow State University, Moscow, GSP-2, Moscow 119992, Russia  
alexberd1983@mail.ru

**Abstract.** The paper presents the module of interactive word sense disambiguation and syntactic ambiguity resolution used within a machine translation system, ETAP-3. The method applied consists in asking the user to identify a word sense, or a syntactic interpretation, whenever the system lacks reliable data to make the choice automatically. In lexical disambiguation, part of man-machine dialogue refers to the analysis phase, while the other part is activated during transfer. For this purpose, entries of the working dictionaries of the system are supplemented with clear diagnostic comments and illustrations that enable the user to choose the most appropriate option and in this way channel the course of system operation.

## 1 Introductory Remarks. ETAP-3 Overview

ETAP-3 is a full-scale rule-based machine translation system that serves Russian-English and English-Russian pairs and has a number of small prototype modules for Russian-German, French-Russian, Russian-Korean, Russian-Spanish and Arabic-English translation. The MT system is developed as part of a multipurpose linguistic processor at the Laboratory of computational linguistics, Institute for Information Transmission Problems in Moscow [1-4]. Other modules of the processor include a parsing tool for deep syntactic tagging of text corpora, a UNL enconverter and deconverter tool, and several smaller-scale components (a module of synonymous paraphrasing of sentences, syntax checker, and a computer-assisted language learning tool).

---

\* This work was supported by a grant (No. 02-06-80085) from the Russian Foundation of Basic Research, whose assistance is gratefully acknowledged.

ETAP-3 is based on the general linguistic framework of the Meaning  $\Leftrightarrow$  Text theory (MTT), proposed by Igor Mel'čuk e.g.[5], complemented by the concept of systematic lexicography and integrated description of language proposed by Jurij Apresjan [6]. However, the classic MTT was somewhat reduced and modified for ETAP-3. In particular, instead of the surface and deep levels of syntactic representation, the system uses a level largely corresponding to surface MTT syntax and a level of normalized syntactic structures in which syntactic relations remain the same but much of national specificity of the source language is removed (see below 1.2 for details).

### 1.1 Morphological Analysis

ETAP-3 processes written texts and translates them sentence by sentence. Every source language sentence is first morphologically analyzed, which means that every word is assigned a deep morphological representation, i.e. the lemma furnished with inflectional characteristics. If a word is morphologically and/or lexically ambiguous, it is assigned a set of morphological representations. Morphological analysis does not take into account any word context, so no lexical or morphological ambiguity is resolved at this stage. The sequence of all morphological representations of the words of a sentence is its morphological structure (MorphS).

The morphological module uses vast morphological dictionaries (the Russian dictionary counts 130,000 lemmas amounting to several million word forms, and the English dictionary counts 85,000 lemmas), and a computationally efficient finite-state software engine. The morphological analyzer is able to parse compound words absent in the dictionary, like English *bioterrorism* or *quasielastic* and Russian *odinnadcatimetrovj* 'eleven-meter' or *neftepererabotka* 'oil processing'.

### 1.2 Parsing

The MorphS of the source sentence is processed by a small pre-syntactic module, which partially resolves lexical and morphological ambiguity using information on close linear context. To give a simple example, if the ambiguous word *lead* is preceded by an article, its verbal interpretation is excluded from further consideration. The partially disambiguated MorphS of the source sentence is sent to the parser – the system's most important and sophisticated part.

The parser transforms the MorphS of the sentence into a dependency tree structure. The tree nodes correspond to the words of the sentence processed, whilst the directed arcs are labeled with names of syntactic relations, or SR. The parsing algorithm creates a dependency tree from the linear MorphS using *s y n t a g m s*, or rules that produce minimal subtrees consisting of two nodes linked by a labeled directed arc. The set of syntagms comprises several hundred rules for each of the two main working languages, written in a specially designed formalism, FORET. Normally, every syntagm describes a specific binary syntactic construction (e.g. nominal subject plus verbal predicate as in *war stinks*, noun plus adjectival modifier, as in *fair play*, numeral plus noun, as in *seven seas*, etc.).

Parser operation consists of several phases. Syntagms create for the given MorphS all possible syntactic links, using all kinds of linguistic and contextual information available. At subsequent phases of parsing extraneous links are eliminated with the help of several filtering mechanisms.

To optimize the parsing process, syntagms are arranged into three types: general syntagms operating on each sentence processed, template syntagms referred to in dictionary entries of restricted word classes, and dictionary syntagms located directly in the entries of syntactically salient words (auxiliaries, conjunctions etc.). This type of rule arrangement is applied in most ETAP-3 phases and modules.

If a sentence is lexically and/or syntactically ambiguous, the parser is able to produce several syntactic structures (SyntS) corresponding to different readings.

An important innovation introduced to parsing theory and practice by ETAP-3 is a mechanism of the detection of the top node that resorts to empirical preference rules. The mechanism consists of several rules (applied after all hypothetical links have been formed) that assign weights to all words of the sentence depending on their likelihood to be the tree top, which is estimated by part-of-speech attribution of each word, its syntactic categorization, presence of sentential markers like conjunctions or connective words, and types of links established. Normally, this mechanism channels the course of parsing in the correct direction and improves the overall performance of the system.

Other recent innovations in the ETAP-3 parser include a system of empirical weights dynamically assigned to the elements of the dependency tree at earlier stages of the parsing process [7] and a module of preference rules based on statistics learned from syntactically annotated corpora [8].

The ready SyntS is sent to the SyntS normalization module that is used to strip the SyntS structure of some of the specific features of the source language. Typical normalization rules merge into single nodes verbal expressions formed with auxiliaries, remove from SyntS strongly governed prepositions and conjunctions, etc. The output of the normalization module is called Normalized Syntactic Structure, or NormS.

### 1.3 Transfer

The transfer proper is performed at the level of NormS, which provides sufficient control of sentence semantics as many of the SSRs are semantically motivated and the nodes carry semantic data inherited from the combinatorial dictionaries of the source language. As a result of the transfer phase operation, the NormS of the source language is replaced by a NormS of the target language, in which all nodes represent the words of the target language and the arcs are labeled with target SSR names. In a way, the NormS is a sort of tradeoff between the two level syntax of MTT and the complexity of the system.

The target NormS is sent to a refinement module which fulfils operations inverse to the ones performed by the normalization module. In particular, it generates analytical verb forms, introduces strongly governed prepositions and conjunctions and ensures the correct word order of the target sentence. The resulting expanded target SyntS is almost ready for the next-to-last phase of translation – syntactic synthesis, which produces the lacking morphological features (as required by agreement or government rules) and prepares ground for the final phase of translation – morphological generation that uses the target morphological dictionary to generate real word forms and produce the target sentence.

#### 1.4 Combinatorial Dictionaries

Combinatorial dictionaries are slightly reduced (they provide no lexicographic definitions) but fully formalized versions of explanatory combinatorial dictionaries (ECD) of Mel'čuk's Meaning  $\Leftrightarrow$  Text theory. The dictionaries are highly reusable; in particular, the Russian combinatorial dictionary is used as the source dictionary in the Russian-to-English translation and as the target dictionary in the opposite direction of translation. For the English combinatorial dictionary, the reverse is true. Currently, each of the two dictionaries contains about 85,000 lexical entries.

An entry of the combinatorial dictionary contains, in addition to the lemma name, information on syntactic and semantic features of the word, its subcategorization frame, a default translation, rules of various types, and values of lexical functions for which the lemma is the keyword. The word's syntactic features characterize its ability (or inability) to participate in specific syntactic constructions. A word can have several syntactic features selected from a total of more than 200 items. Semantic features are needed to check the semantic agreement between the words in a sentence.

## 2 Ambiguity: Persistent Problem

Despite many innovations and rapid advances in both rule-based and statistics-based NLP systems, the disambiguation problem remains a stumbling block for such systems, especially those in which identification of meaning is essential. Recently, much effort has been expended to solve the problem with purely automatic means.

On the one hand, disambiguation techniques have been resorting to more and more sophisticated data supplied in lexical and grammar resources of NLP systems, such as fine-grained constraints on using specific word senses or special rules targeted towards selecting the correct word sense or syntactic interpretation in clearly stated contextual environments. This is what most systems of machine translation are doing, and ETAP-3 has been no exception. Obviously enough, such efforts have their natural limits as they require immense amount of time and labor. Besides, many ambiguous cases cannot in principle be resolved in this way, as they require extralinguistic knowledge far beyond the scope of what can be extracted from texts alone.

On the other hand, significant progress has been achieved in the development of statistical methods designed to disambiguate word senses and trained on large text corpora. This has recently been confirmed by the contributions to the Senseval-3 Workshop held within the framework of the Annual ACL Meeting in Barcelona, as well as papers presented to Coling 2004. Such an approach seems to be more promising; still, characteristically enough, even the most sophisticated statistical techniques (see e.g. [9]) show that the maximum degree of word sense disambiguation achieved on parallel corpora do not exceed 75% – which is impressive but still far from sufficient.

It seems that fully automatic procedures, including the most efficient ones, cannot ensure reliable resolution of linguistic ambiguity.

### 3 Interactive Disambiguation: A Promising Solution

In both approaches listed above, human participation in text processing is confined to preliminary stages (pre-editing) and final stages thereof (post-editing). The approach we have been developing lately takes a different perspective. The human is expected to intervene into text processing in the very heart of the interpretation stage. In machine translation, such a human must know the source language, whereas the command of the target language is not necessary (though of course it can do no harm). In a particular case, such a human may well be the author interested in translating his text into a language with which he is not familiar.

This idea was first put forward some 25 years ago: as evidenced by W. Hutchins [10], American MT systems, ALPS and Weidner in Provo, Utah, used interactive disambiguation of English in early 1980s. Maruyama *et al.* [11] reported the use of the technique for Japanese, progress and elaborated in detail by Christian Boitet and Hervé Blanchon in Grenoble [11-13]. Since then, it has been promoted by a number of research groups in a variety of NLP systems, including 1) the LIDIA dialogue-based machine translation system by the GETA group in Grenoble; 2) the multilingual MT system SYSTRAN; 3) the ALT-J/E system by NTT Communication Science Laboratories of Japan; 4) the UMIST MT system in Manchester, 5) a system of spoken and written translation by the Spoken Translation group in the USA, 6) a system of multilingual search and Internet navigation by DFKI and the University of Saarland in Germany, and many others.

The first system of full-scale interactive disambiguation in NLP for Russian and English was started by the ETAP group in 2002 and has been rapidly progressing since then.

#### 3.1 Lexical Disambiguation

The main idea of the project has been to provide the human expert operating the MT system with clear and simple diagnostic descriptions of ambiguous lexical units that could be viewed at certain phases of text processing. The analysis algorithm has been modified in such a way as to take into account the choices made by the expert and, accordingly, suppress other options that contradict these choices – (possibly, temporarily, in case the choices are incorrect and lead to system failure).

Several points in the algorithm have been specified at which the computer expert opinion is expected: (1) immediately before the parser starts the top node selection; (2) immediately after all syntactic hypotheses generated by syntagms have been checked; (3) immediately before translation options are to be chosen.

As of today, almost 15,000 Russian ambiguous lexical units that share their lemma names (or wordforms) with other lexical units were selected and supplied with diagnostic comments and examples. Information used in these diagnostic comments includes 1) an analytical definition of the word sense, or its important fragment; 2) part of speech tags, which can in case of need be supplemented by simple syntactic features; 3) reference to the word's synonyms and/or opposites. Examples are chosen in such a way as to maximally facilitate word sense identification by the expert.

Optionally, English translation equivalents are supplied for more advanced system users or experts. All information is presented in the respective entries of the Russian combinatorial dictionary. With the help of these comments, many types of lexical ambiguity can be resolved.

At present, the ETAP team is starting a new phase of the project that envisages a similar treatment for the English dictionary. A list of 20,000 ambiguous lexical units of English has been prepared, for which diagnostic comments and examples are being developed.

Importantly, this technique is implemented in a system that strives to obtain all possible analyses of each sentence, rather than choosing just one, even if it is the most probable of all. This approach is motivated by the fact that the system is viewed as a testing ground for a specific theoretical model of language and, insofar as it is possible, must take into account all interpretations that the language allows.

Naturally enough, such an approach narrows the scope of statistical methods applicable in disambiguation. Even though the system has a host of techniques, which can be used to suppress less probable interpretations, we are wary to use them. To be more exact, we want the system to operate in two modes: (a) automatic mode that makes maximum use of probabilistic considerations and discards less probable interpretations at early stages and (b) comprehensive interactive mode, in which the objective is to find any adequate interpretation. In this latter mode of operation, statistic considerations are not discarded altogether but are downplayed a bit.

Another important aspect of our approach is the fact that we treat differently intrinsic ambiguity of the source language and translational ambiguity. This distinction may be disregarded in a system that only serves one pair of languages but it gains in importance in a multilanguage environment. Indeed, some cases of ambiguity must be dealt with regardless of the target language: these can be exemplified by ambiguous English sentences like *he made a general remark*, see below, or Russian sentences like *muzhu izmenjat' nelzja* ('A husband must not be unfaithful' vs. 'One must not be unfaithful to one's husband'). Other ambiguities only arise when we translate something into a particular language. To give a simple example, we should not distinguish between *fish* as animal and *fish* as food when translating from English into Russian but we must do so when translating into Spanish, where *pez* is an animal fish and *pescado* is fish eaten as food. Similarly, we activate ambiguity resolution when translating the Russian adjective *razlichnyj* into English (*different* vs. *various*) and do not activate it when translating into German (*verschieden*).

Since these types of ambiguity are of different character, they are dealt with at different stages of sentence processing: intrinsic ambiguity is treated during analysis whilst translation ambiguity is resolved in transfer. If the distinction is neglected and both types are treated simultaneously, we will have to burden the description of the source language with all ambiguities of all working languages, which is tedious and highly unnatural. Conversely, if we postpone intrinsic disambiguation until transfer, we will miss a good opportunity to discard wrong readings in other parts of the text under treatment.

To the best of our knowledge, ETAP-3 is the only system that clearly distinguishes between these types of ambiguity.

We will now give a couple of examples to illustrate the interactive mode of MT system operation in both directions of translation.

We will start with a short Russian sentence,

- (1) *Soldat tochno vypolnjaj prikazy komandira.*  
 Soldier carried out orders commander

This sentence is highly ambiguous due to the fact that the second word, *tochno*, has no less than four distinctly different senses: TOCHNO1 (an adverb meaning ‘precisely’), TOCHNO2 (an adverb meaning ‘definitely’, ‘by all means’), TOCHNO3 (a comparative conjunction meaning ‘as though’) and TOCHNO4 (a comparative particle meaning ‘like’). All four lexical units are supplied with succinct comments and examples in the combinatorial dictionary. At least three of these senses (TOCHNO1, TOCHNO2 and TOCHNO4) may be in place in (1). If the interactive module of lexical disambiguation of ETAP-3 is on, the human expert will be offered a dialogue window (see Figure 1) and in this way given the chance to choose a word sense which he considers appropriate.

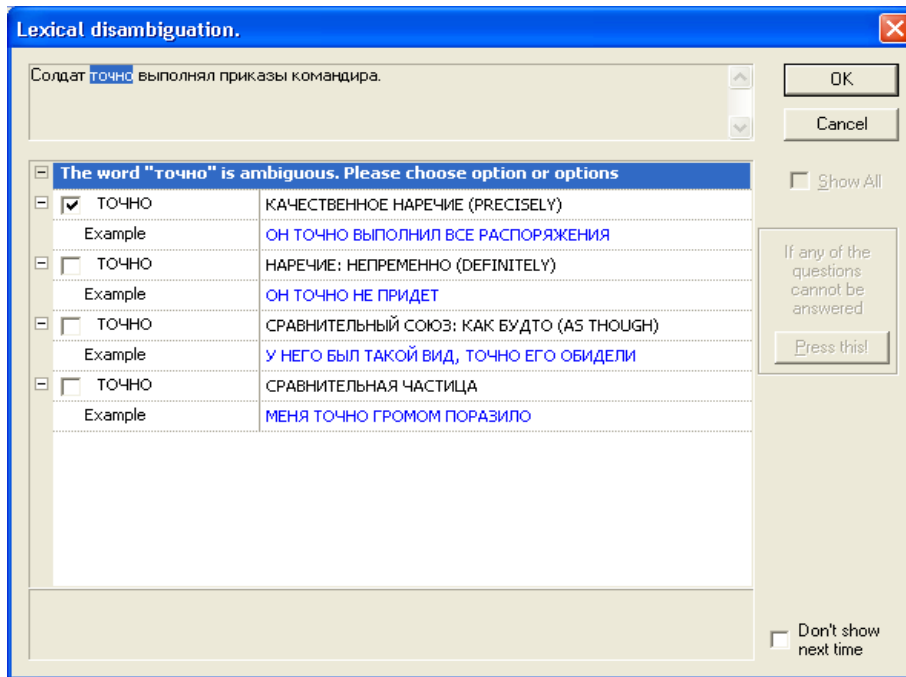


Fig. 1. Dialogue window enabling word sense disambiguation in ETAP-3 for Russian

Using the comments and examples as guidelines, the expert is able to make a reasonable choice and in this way direct the subsequent operation of the system, which eventually lead to the adequate translation. In case of sentence (1), should the first option be chosen, the generated structure will correspond to the interpretation (1a)

*The soldier carried out the commander's orders precisely.* If option 4 is chosen, the obtained syntactic structure will trigger a nontrivial transformation during the transfer phase, which will yield the translation (1b) *It looked like the soldier carried out the commander's orders.* Finally, if the expert highlights option 2, the system will generate the structure corresponding to the interpretation (1c) *The soldier definitely carried out the commander's orders.*

The remaining option 3 (the word *tochno* as a comparative conjunction) is impossible and will be discarded by the system automatically.

As it happens, the fact that the first reading of *tochno* 'precisely' is more frequent in sentences like (1) does not validate the rejection of other readings – because of the principle of multiple interpretation stated above.

Let us now consider an example of lexical disambiguation in English. It is easy to see that sentence

(2) *He made some general remark that everything was fine,*

is ambiguous between (at least) two interpretations: (2a) 'he made some (army) general say that everything was fine' and (2b) 'he made some general observation that everything was fine'. In fully automatic operation, the ETAP-3 MT system yields for sentence (2) two different SyntS (see Fig. 2 and 3), which correspond to the two interpretations.

In Figure 2, *general* is a noun and *remark* is a verb. In Fig. 3, *general* is an adjective and *remark* is a noun. Accordingly, in Fig. 2 the noun *general* serves as the first complement of the verb *make* while the verb *remark* is its second complement (thus creating a complex object construction); cf. labels **1-compl** and **2-compl** on the corresponding links. In Fig. 3, the adjective *general* modifies the noun *remark* (as shown by the **modif** label on the link that connects *general* to *remark*).

Even though ETAP-3 is able to identify this ambiguity, it cannot in the general case automatically decide which of the options is appropriate in a particular context.

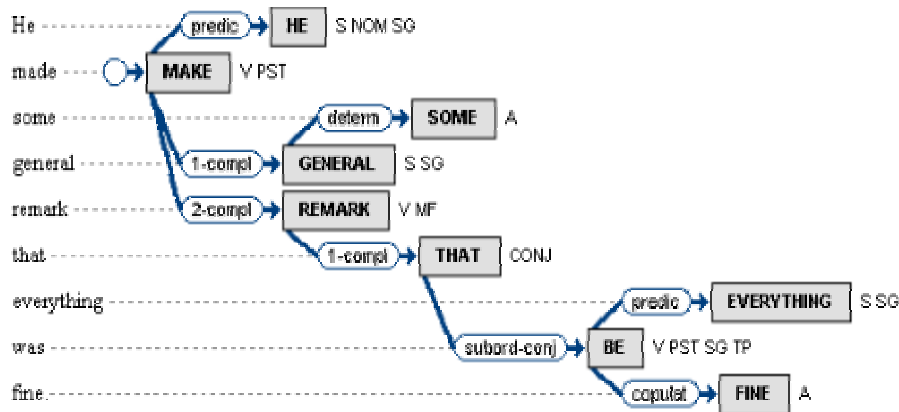


Fig. 2. SyntS for the first reading of sentence (2)



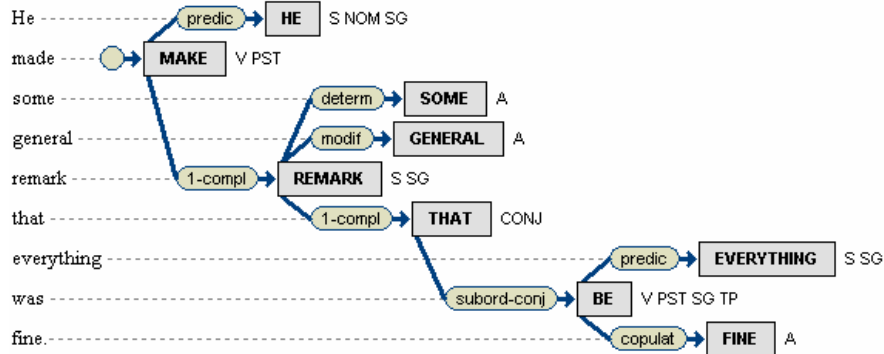


Fig. 3. SyntS for the second reading of sentence (2)

Let us now see what happens if we resort to interactive disambiguation. As in the previous case, the expert will be offered a dialogue window in which he or she has to choose between lexical readings (Fig. 4). If the adjectival reading of *general* and the noun reading of *remark* is chosen, the parser will build the SyntS presented in Fig. 3. Subsequently, the transfer phase will generate the corresponding translation.

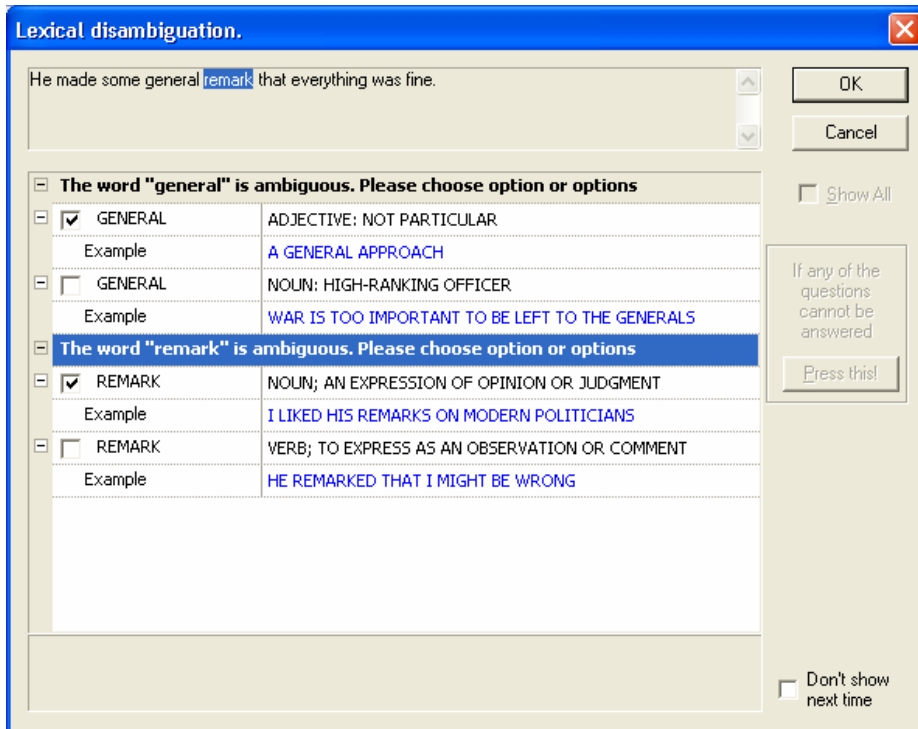


Fig. 4. Dialogue window enabling word sense disambiguation in ETAP-3 for English

Importantly, the module of **lexical** disambiguation helped to solve a rather complicated **syntactic** ambiguity in (2) without actually asking the expert anything about the syntactic structure. Such side effects contribute to broadening the scope of WSD potentials.

It must be emphasized that the lexical disambiguation module can help in far less trivial situations. In the practice of ETAP-3 operation, the system had to translate the following subheading from a recent article on the BBC website:

(3) *AIDS threatens economic collapse.*

For a human, the meaning of sentence (3) is perfectly clear: it says that ‘AIDS endangers (probably, some country) with economic collapse’. In the meantime, the MT system is very likely to understand (3) in an entirely wrong way as ‘AIDS poses a threat to economic collapse’, and, consequently, yield a wrong translation, for the simple reason that the system lacks the resources needed to distinguish the syntactic structure of (3) from that of the sentence

(4) *AIDS threatens economic prosperity.*

Indeed, in order to make sure that (3) is parsed correctly, the system must know that the noun *collapse* instantiates the **instrumental** valency slot of the verb *to threaten* (whatever its sense) and not the **object** slot as in (4). However, to provide adequate word lists for different slots of particular verbs is virtually impossible because, among other things, such lists will inevitably intersect. Cf. ambiguous phrases like *threaten changes*, *threaten a revolution*, or *threaten tax reforms*: unlike *economic collapse*, which is universally viewed as an undesirable event or *economic prosperity*, which is definitely positive, *changes*, *revolutions*, or *tax reforms* may be viewed both positively and negatively. Such an assessment is exactly what a human expert familiar with the text can easily do if asked for a prompt. We do not believe that such cases of intrinsic ambiguity may be successfully solved by statistical methods at all, because this would require collection of data virtually unavailable in any type of linguistic resources (dictionaries or corpora).

### 3.2 Syntactic Disambiguation

It goes without saying that, for the interactive disambiguation system to be really powerful it must also be able to provide the human experts with prompts regarding syntax. This is a difficult task, because average users may readily disambiguate word senses but are normally unprepared to answer questions about the syntax.

The ETAP team is currently investigating possible approaches to the solution of interactive syntactic disambiguation problem. In the meantime, even now the system provides such a module for **an insider**, i.e. a specialist who is well familiar with the particular syntactic module of the ETAP environment. (On a broader scale, this disambiguator module can be used by experts who are specially trained in the system and use it professionally).

Originally, the syntactic disambiguator only offered the user a chance to channel the processing of a sentence by defining whether this was a full verbal sentence or a nominal phrase. This is especially relevant in certain types of English sentences like *Structure changes*  $\approx$  ‘changes of the structure’ vs. ‘arrange the changes’ vs. ‘the

structure is changing' or *Cleaning mechanisms*  $\approx$  'mechanisms for cleaning' vs. 'how to clean mechanisms': such sentences may, with comparable probabilities, be parsed as noun groups or verbal phrases. By determining the type of sentence the user triggered the choice of the top node detection rules.

Currently the module offers a dialogue that enables the human user to choose among syntactic hypotheses, expressed in terms of labeled binary subtrees. The method is especially effective if lexical disambiguation and syntactic disambiguation are applied to a sentence simultaneously.

We will confine ourselves to giving one example, which however is illustrative enough to enable the assessment of the scope and effect of the module as well as the amount of effort needed to use it. The Russian sentence

(5) *Odna iz samyx perspektivnyx oblastej nauchnyx issledovanij – nanotexnologii (texnologii, operirujushchie velichinami porjadka nanometra – nichtozhno maloj velichiny, soppstavimoj s razmerami atoma) – naxoditsja v Rossii poka v zachatochnom sostojanii,*

which has a significant amount of lexical ambiguity and syntactic homonymy, could be processed by ETAP-3 in a fully automatic mode. However, the translation it yielded had some errors and required over 7 minutes of computer time (on a 2 Ghz Pentium 4 computer having 512 Mb RAM).

In striking contrast to that, the use of the lexical and syntactic disambiguation module required one minute of time of an experienced linguist who had to answer about 20 questions, took 1.22 minutes (of which about 1 minute was taken by the linguist), and yielded a far better translation.

(5a) *One of the most promising domains of the scientific investigations - nanotechnologies (technologies, operating with the values of the order of the nanometre - the negligibly small value, comparable to the sizes of the atom), - is in the rudimentary state in Russia so far.*

The results obtained by the development of the two disambiguation modules within the ETAP-3 system so far are encouraging. Future directions of research in this area will focus on the elaboration of means enabling people untrained in the system to use the module of syntactic disambiguation.

## References

1. Apresjan Ju.D., Boguslavskij I.M., Iomdin L.L., Lazurskij A.V. Pertsov, N.V., Sannikov, V.Z., Cinman, L.L. Lingvističeskoe obespečenie sistemy ETAP-2. (The linguistics of the ETAP-2 MT system.) Moskva, Nauka (1989) 295 p. (In Russian.)
2. Apresjan Ju.D., Boguslavskij, I.M., Iomdin, L.L., Lazurskij, A.V., Mitjushin, L.G., Sannikov, V.Z., Cinman, L.L. Lingvističeskij processor dlja slozhnyx informacionnyx sistem. [A linguistic processor for advanced information systems.] Moskva, Nauka (1992) 256 p. (In Russian).
3. Apresjan, Ju.D., Boguslavskij, I.M., Iomdin, L.L., Lazurskij, A.V., Sannikov, V.Z. and Tsinman L.L. Système de traduction automatique {ETAP}. La Traductique. P.Bouillon and A.Clas (eds). Montréal, Les Presses de l'Université de Montréal. (1993).

4. Apresian, Ju.D., Boguslavsky, I.M., Iomdin, L.L., Lazursky, A.V., Sannikov, V.Z., Sizov, V.G., Tsinman, L.L. ETAP-3 Linguistic Processor: a Full-Fledged NLP Implementation of the MTT. // MTT 2003, First International Conference on Meaning – Text Theory. Paris, École Normale Supérieure, Paris, June 16–18, 2003, pp. 279-288.
5. Mel'čuk I.A. Opyt teorii lingvisticheskix modelej klassa "Smysl - Tekst". [The theory of linguistic models of the Meaning – Text Type]. Moscow, Nauka. (1974) (In Russian).
6. Apresjan, Ju. D. Systematic Lexicography. Oxford University Press (2000) XVIII p., 304 p.
7. Iomdin, L.L., Sizov, V.Z., Tsinman, L.L. Utilisation des poids empiriques dans l'analyse syntaxique: une application en Traduction Automatique. META, 47. (3). (2002) 351-358
8. Boguslavskij I.M., Iomdin L.L., Sizov, V.G., Chardin, I.S. Ispolzovanie razmechennogo korpusa tekstov pri avtomaticheskom sintaksicheskom analize. (Using a syntactically tagged corpus of texts in automatic syntactic analysis).// Proceedings of the International Conference "Cognitive Modeling in Linguistics 2003. Varna (2003). 39-48.
9. Tufis D., Ion, R., Ide, N. Fine-Grained Word Sense Disambiguation Based on Parallel Corpora, Word Alignment, Word Clustering and Aligned Wordnets. // Proceedings of the 20th International Conference on Computational Linguistics, Geneva, August 23–27, 2004, pp. 1312-1318.
10. Hutchins W. Machine translation: past, present, future. Ellis Horwood, Chichester (1986).
11. Maruyama H., Watanabe, H., and Ogino, S. An interactive Japanese parser for machine translation. // Karlgren, H., ed. Proceedings of the 13th International Conference on Computational Linguistics, v. 2. Helsinki. (1990). 257-62
12. Boitet, C. & Blanchon, H. Multilingual Dialogue-Based MT for monolingual authors: the LIDIA project and a first mockup. // Machine Translation, 9/2 (1994), 99-132.
13. Blanchon, H. An Interactive Disambiguation Module for English Natural Language Utterances. // Proceedings of NLPRS'95. (Seoul, Dec 4-7, 1995), vol. 2/2: 550-555.
14. Goodman, K. and Nirenburg, S. (ed.). The KBMT Project: A case study in knowledge-based machine translation. Morgan Kaufmann Publishers. San Mateo, California. (1991). 330 p.