

Recent developments in Russian–French machine translation at Grenoble

CHRISTIAN BOITET and NIKOLAI NEDOBEJKINE

Abstract

This paper describes an application of the researches pursued at GETA (Groupe d'Études pour la Traduction Automatique) on the mechanical translation from scientific Russian texts into French. In the first part, we briefly present ARIANE-78, a general computer system designed to be used by linguists for writing translation systems. The system includes several components, each of them providing a metalanguage in which linguistic data (such as grammatical variables, classes, dictionaries and grammars) are expressed, a compiler and general execution programs. An interactive monitor manages all files and offers a transparent interface to the user.

In the second part, we present in turn each main step of the translation process, as developed for the Russian–French application: morphological analysis, multilevel analysis, lexical transfer, structural transfer, syntactic generation, morphological generation. For each such step, we present first the underlying algorithmic model and computer system, then the linguistic data and strategies, and an example of execution over a small test text.

Introduction

This paper is an attempt to present the computer models and linguistic strategies used in the current version of the Russian–French translation system developed at GETA, within the framework of several other applications which have been developed in a parallel way, using the same computer system. This computer system, called ARIANE-78, offers to linguists not trained in programming an interactive environment, together with specialized metalanguages in which they write linguistic data and procedures (essentially, dictionaries and grammars) used to build translation systems. In ARIANE-78, translation of a text occurs in six steps: morphological analysis, multilevel analysis, lexical transfer, structural

transfer, syntactic generation, morphological generation. To each of these steps there is a corresponding computer model (non-deterministic finite-state string to tree transducer, tree to tree transducer, ...), a metalanguage, a compiler and execution programs. The units of translation are not sentences, but rather one or several paragraphs, so that the context usable, for instance to resolve anaphors, is larger than in other second-generation systems.

As ARIANE-78 is independent of any particular application, we begin by presenting its main features in Part I. Some of them are standard in second-generation systems, while others are original. Among these, we stress the multilingual aspect of the system, which is quite unique, the very powerful control structures embodied in the various computer models (non-determinism, parallelism, heuristic programming), and its interactive data-base facility.

In the second and more voluminous part, we successively describe each step of this Russian-French application. First we present the underlying computer model (there are only four as the second, third and fourth step use the same one), then the organization of the linguistic data. A small text is used throughout this paper as a standard example. Appendix 1 contains several examples of unrevised machine translations of larger texts, with some comments. Appendix 2 gives the variables used in the different steps, with some comments.

1. CURRENT GETA TRANSLATION SYSTEM

The computer system ARIANE-78, together with appropriate linguistic data, constitutes a multilingual automatized translation system.

1. General principles

The system is a rather sophisticated second-generation system. It relies on classical as well as more original methods.

1.1. Classical second-generation methods

Intermediate structures. The process of translation of a text from a 'source' language into a 'target' language is split up into THREE MAIN LOGICAL STEPS, as illustrated by Figure 1 below: ANALYSIS, TRANSFER and GENERATION. The output of the analysis is a STRUCTURAL

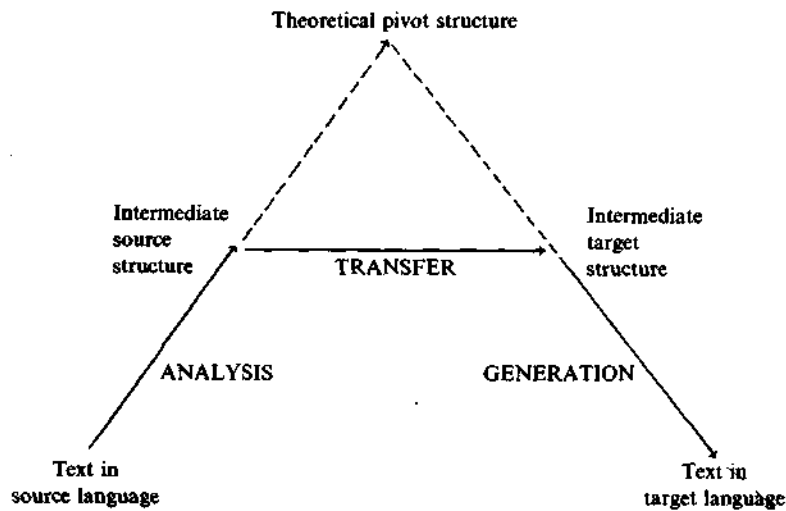


Figure 1.

DESCRIPTOR of the input text, which is transformed into an equivalent structural descriptor in the target language by the transfer phase. This target structural descriptor is then transformed into the output text by the generation phase. Essential to our concept is the fact that analysis is performed independently of the target language(s). The 'deeper' the analysis, the shorter the distance between the two structural descriptors. Ideally, one could imagine a 'pivotal' level, at which they would be the same.

In the past, Prof. Vauquois' team tried a slightly different idea (Vauquois, 1975), namely to use a 'hybrid' (Shaumjan) pivot language, where the lexical units are taken from a natural language, so that the transfer phase is reduced to a lexical transfer, without any structural change. As it is not always possible, or even desirable, to reach this very abstract level, one may choose not to go all the way up the mountain but to stop somewhere in the middle. This is why we call our structural descriptors 'INTERMEDIATE STRUCTURES'. Note that ARIANE-78 does not impose any particular transfer level, and both the pivot and the superficial extremes are possible, and in fact the linguistic teams have agreed on 'multilevel' intermediate structures which contain very deep as well as low level types of information, ranging from logical relations to traces (see details in 1.2. below).

Separation of programs and linguistic data. The second classical method is to offer metalanguages, in order to keep the particular linguistic data

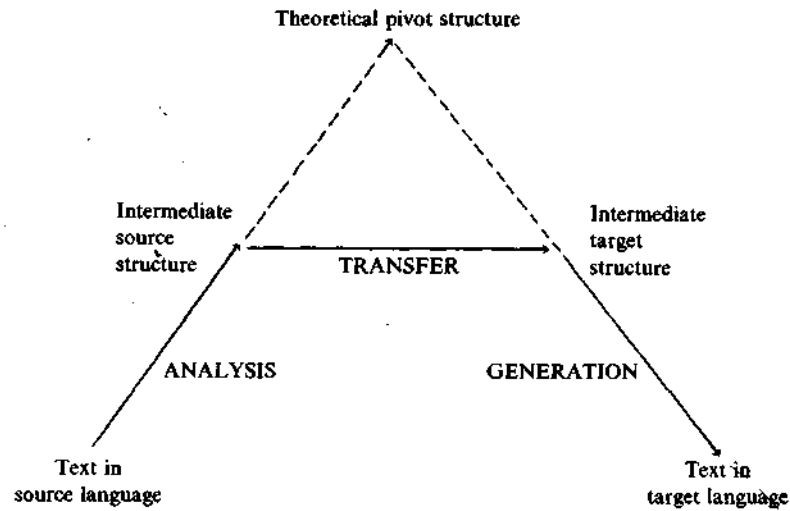


Figure 1.

DESCRIPTOR of the input text, which is transformed into an equivalent structural descriptor in the target language by the transfer phase. This target structural descriptor is then transformed into the output text by the generation phase. Essential to our concept is the fact that analysis is performed independently of the target language(s). The 'deeper' the analysis, the shorter the distance between the two structural descriptors. Ideally, one could imagine a 'pivotal' level, at which they would be the same.

In the past, Prof. Vauquois' team tried a slightly different idea (Vauquois, 1975), namely to use a 'hybrid' (Shaumjan) pivot language, where the lexical units are taken from a natural language, so that the transfer phase is reduced to a lexical transfer, without any structural change. As it is not always possible, or even desirable, to reach this very abstract level, one may choose not to go all the way up the mountain but to stop somewhere in the middle. This is why we call our structural descriptors 'INTERMEDIATE STRUCTURES'. Note that ARIANE-78 does not impose any particular transfer level, and both the pivot and the superficial extremes are possible, and in fact the linguistic teams have agreed on 'multilevel' intermediate structures which contain very deep as well as low level types of information, ranging from logical relations to traces (see details in 1.2. below).

Separation of programs and linguistic data. The second classical method is to offer metalanguages, in order to keep the particular linguistic data

(grammars, dictionaries) separated from the programs. For instance, dictionary look-up is a standard function, which should not be modified in any way when a new language is introduced in the system. This separation also corresponds to a division of work and enhances transparency: dictionary look-up may be optimized by the programmers without the linguistic users ever being aware of it. The same goes for more complex functions, like pattern-matching in tree-manipulating systems. In these metalanguages, linguists work directly with familiar concepts, like grammatical variables, classes, dictionaries and grammars: grammar rules are rules of some formal model (context free, context sensitive, transduction rules). That is, one may also consider such metalanguages as very high level algorithmic languages offering complex data types and associated operators. Although this principle of separation has been criticized as imposing too much 'rigidity' on the users, critics have failed to understand that this is only the case when the metalanguages are inadequate. A good comparison may be found in classical programming, where for example, in exactly the same sense, the compiler and run-time package of PL/I are separated from programs written in PL/I.

Semantics by features. The third classical principle touches SEMANTICS. In second-generation MT systems, semantics may only be expressed by the use of FEATURES (concrete, abstract, countable, ...), which are exactly like grammatical features. The theoretical framework is that of a formal language, with a syntax describing the combination rules of the language units. There is no direct way, for instance, to relate two lexical units. In order for this to be possible, there should be a (formalized) domain, possibly represented as a thesaurus, and rules of interpretation. However, this limitation may be partially overcome in ARIANE-78's lexical transfer step. Note also that semantic features may be extremely refined for some limited universe, and give surprisingly good results (TAUM-METEO, in operation since 1975).

1.2. *Methods unique to GETA's system*

These are discussed below under the same headings as above.

Intermediate structures. In ARIANE-78, we split up each of the three main phases into two steps. This is essentially for algorithmic as well as for linguistic reasons. Morphological analysis, lexical transfer and morphological generation are undoubtedly very much simpler than the other steps, and it seemed reasonable and linguistically motivated to keep them

separate and to use simpler algorithmic models to realize them. However, this could not be done in other environments, for example if the input were expected to be very noisy (oral input).

ARIANE-78 uses a unique kind of data-structure to represent the unit of translation from morphological analysis to morphological generation, namely a COMPLEX LABELLED TREE STRUCTURE: each node of such a tree bears a value for each of the 'grammatical variables' used in the current step (see 3 below).

GETA'S system is MULTILINGUAL by design: an analysis cannot explicitly use information from the target language, and generation is likewise independent of the source language. Moreover, for one particular user, ARIANE-78 ensures the coherence of the linguistic data written to construct a multilingual application.

Computer environment. The principle of separation of programs and linguistic data is strictly observed in our system. An additional feature is to propose SEVERAL ALGORITHMIC MODELS designed to be of maximal adequacy and generality as well as of minimal computational complexity.

Functions of an integrated MT system include preparation of the linguistic data, management of the corpora and execution of the linguistic data over texts. ARIANE-78 provides a CONVERSATIONAL ENVIRONMENT for these functions, hiding implementation chores from the user. It also includes a SPECIALIZED DATA-BASE MANAGEMENT SYSTEM for texts and linguistic files.

Semantics. Semantic features may be declared as normal grammatical features in each step. At lexical transfer, the linguist may relate several source and target lexical units, these relations being elaborated in the succeeding structural transfer phase. This is, however, certainly not sufficient to call the system 'third generation'. Because of this, we are perplexed that some people refer to Systran as 'third generation'.

2. Organization of the translation process

2.1. Overall schema

The schema in Figure 2 shows the different steps of the translation process. The components of ARIANE-78 implementing the four different algorithmic models appear within circles: they are linked by double lines to the rectangles corresponding to the linguistic data written in the

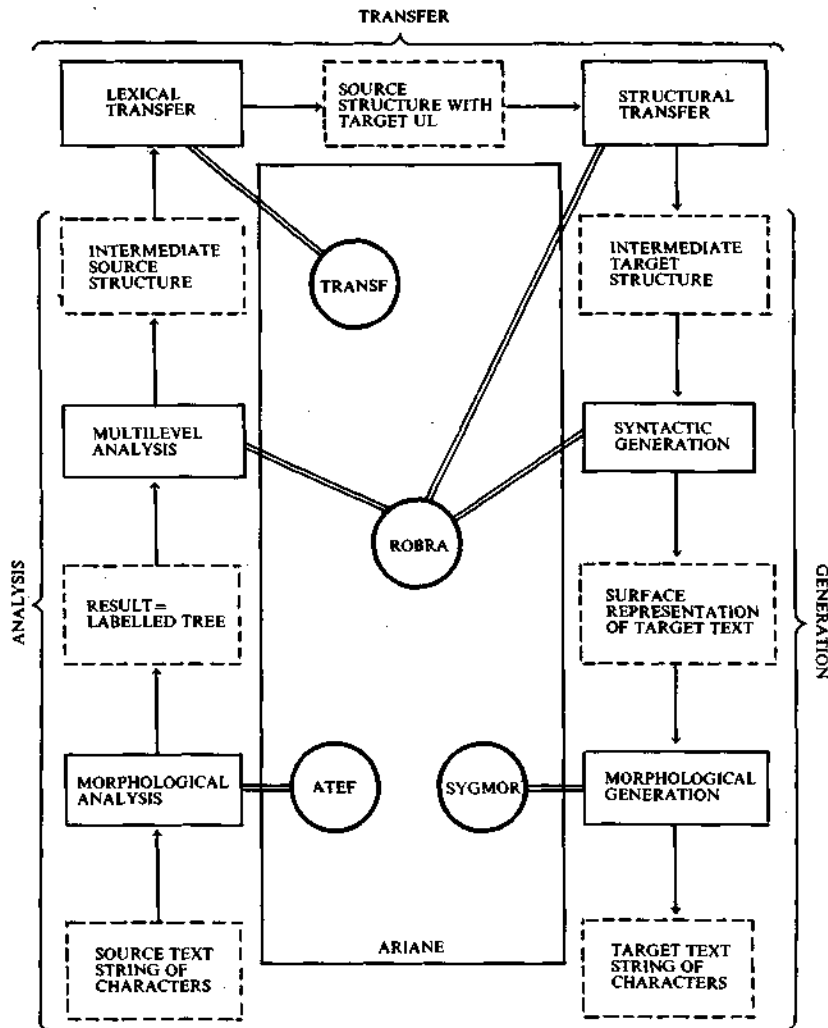


Figure 2.

associated metalanguage for the indicated step. Simple arrows show the flow of control. UL means 'lexical unit'.

2.2. Organization of a step

In each step, the linguistic data may be of four different kinds:
 — grammatical variables (like gender, number, semantic type)

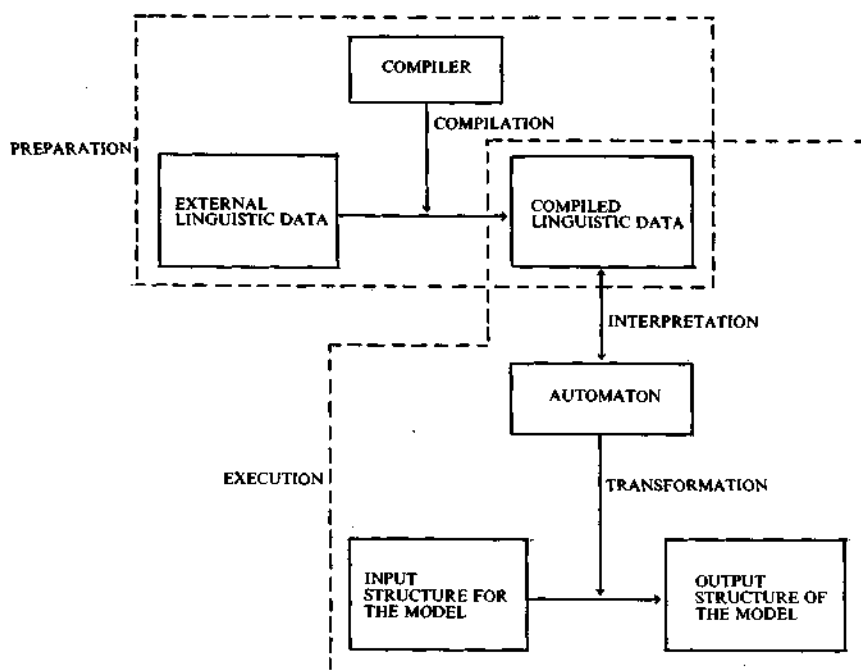


Figure 3.

- classes, describing useful combinations of values of variables
- dictionaries
- grammars, containing the rules and the strategy for using them.

They are expressed in a metalanguage. Their syntax and coherency is first checked by the corresponding compiler which then generates a compact intermediate code. At run-time, this code is interpreted¹ by standard 'execution programs'. This organization is shown in Figure 3.

This approach separates the linguistic and algorithmic problems, and makes debugging and maintenance much easier.

2.3. Implementation

The complete system runs on IBM machines, Series 370 and above, under VM/CMS. Each user's workspace is implemented as a virtual machine with a virtual memory of one to two megabytes and a disk space varying with the size of the application. ARIANE-78 resides on its own virtual machine whose disk space is shared by the different users.

The conversational monitor integrates the following functions:

- preparation of the linguistic data (creation, modification, compilation, listing, sorting).
- preparation of the texts (same functions, plus specific ones such as the listing of words and the display of structures). They are organized by corpora. For each corpus, there is a hierarchical list of separators giving a method of structuring the texts (e.g. in chapters, sections, paragraphs and sentences).
- execution of (part of) the translation process over a text. The execution of each step is done according to user-provided parameters. In particular, tracing is quite sophisticated. Interactive modes of execution exist in ATEF and ROBRA. Intermediate results are stored and retrieved automatically by the system.

3. Terminology relative to the representation of the information

Let us first give some definitions to fix our terminology, then an example and some ideas about the usage of these facilities.

3.1. Formal units of information

A VARIABLE is defined by a name and a list of elementary values. The set of possible values is made up of:

- the set of elements of the list and a 'null' value, if the variable is 'exclusive'.
- the set of subsets of the above set, if the variable is 'not-exclusive'.
- the set of the relative integers of absolute value less than or equal to the single element of the list, for an 'arithmetic' variable.
- the set of lexical units defined in the dictionaries and in the grammars, for the special variable UL.

For example, to define a variable of gender, to have three values (male, female, neuter), we write GENDER:=(MALE, FEMALE, NEUTER).

A MASK OF VARIABLES is a combination of values of the variables used (and declared to the system). The labels borne by the tree structures which are manipulated will always be such masks of variables; moreover, the set of variables may change from one step to the next. Contrary to a too widespread habit (in linguistics), the labels of the nodes of trees are therefore complex. This allows us to avoid the introduction of unnecessary problems like artificial discontinuities and to separate the geometrical properties of the structures from their interpretation.

A **FORMAT**, or **CLASS**, is a constant mask of variables which has been given a name and which can be used as reference in dictionaries and grammars.

A **FORM**, or **OCCURRENCE**, is a succession of non-blank characters bracketed by two blanks.

A **LEXICAL UNIT** is a value of the variable **UL**, predefined and exclusive. The lexical units are introduced in the dictionaries and the grammars, rather than as a declared list of values. In this sense, **UL** may be called 'potential'.

A **LABELLED TREE STRUCTURE** is a tree structure where each node carries a mask of variables.

3.2. Example

Declaration of variables:

CAT := (ART, NOM, VB, ADJ). ** Syntactic category.
 K := (GN, GP, GV). ** Syntagmatic class.
 NB := (S, P). ** Number.
 FS := (DES, EPIT, GOV, SUJ, OBJ). ** Syntactic function.

The structure (from a sentence like 'the linear systems are such that ...').

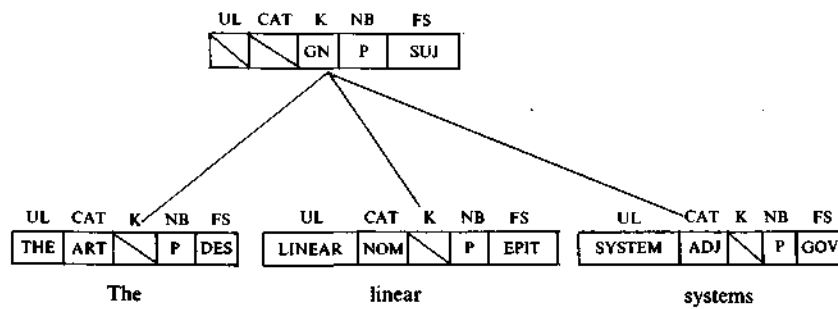


Figure 4.

Masks of variables are implemented as fixed-length records of 32 bytes in **ARIANE-78**. Each variable will be mapped by the compiler into a field of such a record.

3.3. Usage and interpretation

The type of structure supported by the system is quite general and does not force any particular representation technique (like dependency or by immediate constituents) on the user. The following principles of usage and

interpretation have been worked out by Prof. Vauquois and are being used in several applications written in ARIANE-78.

(1) The graph of the structure only gives a bracketing of the text. Graphical dependency does not imply any linguistic dependency relation.

(2) The information appearing on the nodes allows interpretations of the structure at three main levels, corresponding to:

- syntactic and syntagmatic classes
- syntactic functions
- logical relations (akin to deep cases).

It is not always possible to compute the values of the 'deepest' variables everywhere in a unit of translation. This technique makes it possible, in case of failure to compute the logical relations for some part of the translation unit, to fall back on the syntactic functions, but still to use the logical relations elsewhere. This 'safety net' strategy would not be possible if different graphs corresponded to different levels of analysis.

II. AN APPLICATION TO RUSSIAN-FRENCH TRANSLATION

Introduction

The following translation has been produced automatically. It will serve as our standard example throughout this section.

Input text:

SFORMULIROVAN PRINCIP, S POMOTHQYU KOTOROGO OPREDELYAETSYA KRITERIJ, PRIGODNYU DLYA NELINEJNOJ TERMODINAMIKHESKOJ SISTEMYI.

Output text:

ON A FORMULE LE PRINCIPE A L'AIDE DUQUEL ON DEFINIT LE CRITERE UTILE POUR LE SYSTEME THERMODYNAMIQUE NON-LINEAIRE.

English equivalent:

'The principle with whose help the criterion useful for the non-linear thermodynamic system is defined, has been formulated'.

Although we have taken a one-sentence text for the sake of brevity, we stress that units of translation are not sentences. As a matter of fact, sentence boundaries are detected in the morphological analysis step. Appendix 1 gives longer examples.

As cyrillic characters are not included in the EBCDIC character set and are unavailable on standard equipment, we use the latin transcription

given below. As there are more letters in the cyrillic alphabet than in the latin one, some cyrillic letters are mapped onto two latin letters. No ambiguity arises, as 'H' and 'Y', which appear in these combinations, never appear alone.

А	Б	В	Г	Д	Е	Ж	З	И	Й	К	Л	М	Н	О	П
A	B	V	G	D	E	ZH	Z	I	J	K	L	M	N	O	P
Р	С	Т	У	Ф	Х	Ц	Ч	Ш	Щ	Ъ	Ы	Ь	Э	Ю	Я
R	S	T	U	F	X	C	KH	SH	TH	W	YI	Q	YE	YU	YA

Figure 5.

Note that we may, and in fact do, have several versions of the analysis or generation of a language, and hence of transfer between two languages. Each such version is identified in ARIANE-78 by a three-character code: we may have RU1 and RU2 for analysis, FR1 and FR2 for generation. The following presentation applies to the RUS-FRA pair, which has been now kept stable (since June, 1980) for reference purposes, while other versions are evolving. It has been developed on a rather broad spectrum of technical or scientific texts, ranging from small abstracts to long papers, and concerning aeronautics, spatial sciences, thermodynamics and nuclear physics. What follows does not try to give a detailed documentation, but rather to show the essential ideas of the linguistic strategies.

1. Morphological analysis

1.1. The algorithmic component: ATEF

ATEF (Analyse de Textes en Etats Finis) is based on the non-deterministic finite state transducer model. Its external (linguistic) data comprise:

- declarations of variables and formats (classes)
- dictionaries (in this implementation, a maximum of six, plus a dictionary of fixed idioms), in which each item contains a segment ('morph') as access key, and information made up of two formats and possibly a lexical unit.
- a grammar, in which each rule comprises a list of 'calling formats', conditions and actions.

The system successively analyses each form in the text, examining *a priori* all possible analyses. Each stage of a particular analysis consists of cutting out a segment from 'what is left of' the form to be analyzed and in

applying one of the rules referenced by the 'morphological' format associated with this segment in the corresponding item retrieved from a dictionary. For strategic purposes, variables and formats in this system are divided into 'morphological' and 'syntactic'. This division is, however, somewhat misleading, since linguists may use both codes for any kind of information, including semantic features.

The conditions may relate to the results of the analysis of the four preceding forms, to the accessible strings and to the partial results stored by this analysis. It is also possible to store a condition on the result(s) of the analysis of the following form. A particular condition consists in giving a list of 'sub-rules' and requiring that at least one of them applies to the result of the current rule.

There are three types of actions: assignment of values to the system-defined mask 'C', containing the current result of the analysis, transformations on the part of the form which has not yet been segmented, and special functions. These functions make it possible to:

- control the built-in back-track search by eliminating certain possibilities and by opening and closing dictionaries (this is formally done by assignments of the special variable DICT).
- store the current result (useful for compound words).
- create new lexical units from the form processed.
- decide that a sentence boundary has been reached.

In the case of an unrecognized form ('unknown word'), the system restarts the analysis by connecting the form to a special format, which in particular must call the obligatory 'unknown word' rule MOTINC. As this rule may invoke subrules, and as this format may call other rules, this opens the door to elaborate strategies. Moreover, in interactive mode, the system will stop when encountering an unknown word and ask the user if he wants to change the input (e.g. spelling error) or to index tentative new items in the dictionaries.

Lastly, note that the direction of analysis of each form (L-R or R-L) is a parameter fixed by the linguist at compilation time.

Formally, the output of this system is a graph whose nodes are the masks (or groups of masks for compound words) found and where the vertices indicate the compatibility of the analyses with respect to the grammar. Different presentations of this output, tree-diagrams or otherwise, are possible (Q-graphs, labelled tree structures with or without 'homophrases').

1.2. Morphological analysis of Russian

The 'RUS' application uses four dictionaries, two for the stems, one for the endings (giving no lexical units) and one for the fixed idioms.

The analysis of each form occurs from left to right. The only prefixes segmented are 'NE' and 'full' prefixes like 'GIPER-', 'MIKRO-', 'MAKRO-' ... Endings comprise what is classically divided into affixes and flexional endings. This doesn't by any means mean that the corresponding linguistic information is ignored or lost, it is just one of several possible strategies for using ATEF for morphological analysis of Russian, and this one is believed to be reasonable, as measured in terms of coding effort (indexing of dictionaries) and computational efficiency. Further very precise information may be found in the indexing manual (Axtmeyer *et al.*, 1979).

Let us now give an example on the form 'SFORMULIROVAN' (has been formulated).

Examples taken from the dictionaries:

Stems	LINE	==B8ENC	(QUALITE, LINEEN).
	NE	==XPFNE	(NEGAT, NE-).
	SFORMULIR	==B1YDP	(V2, FORMULIROVATQ).
	SISTEM	==B5OJA	(XCOPO, SISTEMA).
Endings	20I	==DS1H	(Q4S).
	80OJ	==DA1	(Q4567F).
	99N	==DVN	(QMAS).
Fixed idioms	DLYA TOGO KHTOBYI	==YCDS2	(FS00, DLYATOGOKHTOBYI).
	MOZHET BYITQ	==XCIRC	(FS00, MOZHETBYITQ).

Figure 6.

Example of rules (see Figure 7).

VB1B and VP2 are the names of two rules. B1YDP, B1YDI, ..., DVM are 'calling formats'. Following the '=', we find the assignments on 'C', the special functions ('-ARRET-' will stop the analysis if application of VP2 leads to a solution) and the control over the dictionaries. Then follow the condition of the rule, the 'string transformation' part, and (not present

VB1B: BIYDP — BIYDI — BIYDD == DICT(C):=3, VAR(C):=VAR(A),
 CS(C):=CS0, ANM(C):=ANM0 /
 SCHAINÉ (A, 0, 1) -E- 'U' -ØU-SCHAINÉ (A, 0, 3) -E- 'ØVA' /
 TCHAINÉ (0, 'U', '91'), TCHAINÉ (0, 'ØVA', '99').

VP2: DVN — DVM == T(C):=T(A), MD(C):=PRT, K(C):=AQ,
 RF(C):=PF, FN(C):=FN0, A(C):=A(C) -I- A(A),
 -ARRET-, DICT(C):=3, FM(C):=FØC, G(C):=G(A),
 N(C):=N(A), P(C):=P(A) /
 SCHAINÉ (A, 0, 1) -E- ' ' -ET- RF(C) -NE-R-ET- CP1(C) -INC-ACC.

Figure 7.

here) the conditions for the following form and the sub-rules. The meaning of all the variables used in RUS-FRA is given in Appendix 2. Some of them will also be explained below.

Example of a trace.

OCCURRENCE: SFORMULIROVAN

MORPHE: SFORMULIR REGLE: VB1B FORMATS: BIYDP V2

ETAT COURANT: SFORMULIROVAN: UL(FORMULIROVATQ), A(P),
 FN(W), CP1(ACC).

RESTE: 99N

MORPHE: 99N REGLE: VP2 FORMATS: DVN QMAS

→DECOUPAGE SFORMULIR 99N

SFORMULIROVAN: UL(FORMULIROVATQ), K(AQ), MD(PRT), P(1, 2, 3),
 T(PAS), FM(FØC), G(M), A(P), N(S), RF(PF), CP1(ACC).

Figure 8.

'SFORMULIROVAN' is the input form. 'SFORMULIR' is a segment, and the corresponding information fills the system-defined 'argument' mask 'A', while 'C' is null. VB1B is applied. It transfers all variables from A to C (VAR(C):=VAR(A)) and opens dictionary three (endings). The following 'ØVA' is changed to '99'. This is a strategic trick to speed up the analysis. The partial result says that the aspect is perfective, the last consonant is hard and this UL has a strong first valency.

The rest to be analysed is '99N'. VP2 is applied. Now, A contains the information associated with '99N'. VP2 transfers tense T, gender G and person P from A to C, and directly assigns mode MD, syntagmatic

category K, as well as RF, FN and FM (short form of the adjective). Aspect A is assigned the intersection of the values given by the stem and the ending. See Figure 9.

2. Multilevel analysis

Introduction

This is the longest part of the paper, because this step is the most difficult, linguistically, and because we have to present ROBRA, the algorithmic component used for this step as well as for structural transfer and syntactic generation. The three other steps each use a specific model (ATEF, TRANSF, SYGMOR) where the external data comprises one or more dictionaries, as opposed to ROBRA, where the external data consist of:

- declarations of variables and formats
- a transformational system, comprising
 - condition and assignment procedures
 - transformational rules
 - transformational grammars
 - a control graph specifying the flow of control.

The input to this step is the flat tree resulting from the morphological analysis. Its output is the source intermediate structure which has been discussed above. The idea is to gradually transform the input into the desired output by applying local transformations in the object tree. These transformations are expressed by rules, and the rules are grouped in grammars, so that, roughly speaking, each grammar corresponds to some coherent linguistic process. This enforces modularity and is an answer to the recurring problem of unmanageability of large grammars for natural systems written in inadequate models (like context-free).

2.1. The algorithmic component: ROBRA

2.1.1. Justification of the type of the model. ROBRA has been presented elsewhere (Chauché, 1974; Boitet *et al.*, 1978) in more formal and more detailed ways. As a computer system, it implements an abstract TREE-TRANSDUCTION MODEL (TTM). Our principle to use transducers rather than analysers can be supported by the following arguments.

First, an analyser accepts or rejects its input. The structure produced by

SUMMET 1 : UL(ULTXTI).
 SUMMET 2 : UL(ULTRA).
 SUMMET 3 : UL(ULOC).
 SUMMET 4 : *SQRMULIROVAN: UL(FORMULIROVATQ),K(AQ),M(PRT),P(1,2,3),T(PAS),FM(FOC),G(M),A(P),NIS),RF(PF),CPL(ACC).
 SUMMET 5 : UL(ULOC).
 SUMMET 6 : PRINCP: UL(PRINCIP),K(NM),C(1,2),P(2),G(M),NIS),NAT(OBJ),PG(6DB),ANN(INA).
 SUMMET 7 : UL(ULOC).
 SUMMET 8 : UL(,),K(VG).
 SUMMET 9 : UL(ULOC).
 SUMMET 10 : *ULPRIPOMUTHI),K(PP),CS(ACT),CPL(GFM),ANN(INA).
 SUMMET 11 : UL(ULOC).
 SUMMET 12 : KOTOROG: UL(KOTORYEJJ),K(AQ),CE(,4),FM(FOL),G(M),N),NIS),CS(DF),VL(IT).
 SUMMET 13 : UL(ULOC).
 SUMMET 14 : *PREDELYAETSYA: UL(OPREDELITQ),K(VB),M(VRB),P(3),T(PRE),S(F,M,N),A(I),NIS),RF(R),CPL(ACC),PS(4IZ),
 C(P2IPP).
 SUMMET 15 : UL(ULOC).
 SUMMET 16 : KRITERIJ: UL(KRITERIJ),K(NM),C(1,2),P(3),G(M),NIS),CS(ACT),CPL(DAT),ANN(INA).
 SUMMET 17 : UL(ULOC).
 SUMMET 18 : UL(,),K(VG).
 SUMMET 19 : UL(ULOC).
 SUMMET 20 : PRIGOVYI: UL(PRIGODFM),K(AQ),C(1,2),FM(FOL),G(M),NIS),CS(ACT),CPL(PP),PG(5K),ANN(INA).
 SUMMET 21 : UL(ULOC).
 SUMMET 22 : OLYA: UL(OLYA),K(PP),C(4),PG(4OLA).
 SUMMET 23 : UL(ULOC).
 SUMMET 24 : UL(FORMCP).
 SUMMET 25 : NELSUNJ: UL(NE-),K(IV),CS(TDF),NG(ME).
 SUMMET 26 : NELSUNJ: UL(LINEEN),K(AQ),C(4,5,6,7),FM(FOL),G(F),NIS),CS(QAL).
 SUMMET 27 : UL(ULOC).
 SUMMET 28 : TERMODINAMIKHESKOJ: UL(TERMODINAMIKHESKOJ),K(AQ),C(4,5,6,7),FM(FOL),G(F),NIS),CS(PLF).
 SUMMET 29 : UL(ULOC).
 SUMMET 30 : SISTEMY: UL(SISTEM),K(MH),C(4),P(3),G(F),N(S),NAT(OBJ),PG(5PD),ANN(INA).
 SUMMET 31 : SISTEMY: UL(SYSTEM),K(MH),C(1,2),P(3),G(F),N(PI),NAT(OBJ),PG(5PD),ANN(INA).
 SUMMET 32 : UL(ULOC).
 SUMMET 33 : UL(,),K(VG).

Figure 9.

an analyser is just a 'byproduct' of a SUCCESSFUL analysis, and nothing is produced in the case of a failure. But it is an unfortunate fact of life that grammars are never perfect and input texts often incorrect (with respect to them or even to normal usage). A transducer may process imperfect inputs and will give a result in any case. Moreover, its output is not directly related to the way it has been produced.

As ROBRA is a tree to tree transducer, the output structure is homogeneous with the input structure and may be processed by the same type of automaton. This allows composition of transformational grammars to be used for MODULARITY. Structured 'linguistic programming' is thus encouraged by the nature of the model and by the structure of its metalanguage.

ROBRA may be viewed as a very high-level algorithmic language, in the sense that its data types and control structures are very high-level. The data types are variables, masks of variables and labelled trees, with the associated operators (comparison and assignment for the first two, transformation by one or several rules for the third). The control structures include conditional assignments (of variables), parallelism (of application of rules in a grammar) and non-determinism (in the control graph). Care has been taken to limit undecidability, which may be guaranteed by syntactic checks performed at compile time. Rules may recursively call (sub) grammars or transformational subsystems.

2.1.2. Transformational rules (TR). A TR is made up of a name, a left part (the SCHEMA) and a right part (the IMAGE). It may be understood as a 'production rule', the schema being the condition and the image the action.

The schema expresses the possible subtrees which the rule may transform. It includes a GEOMETRIC PATTERN and CONDITIONS ON THE VARIABLES appearing on the nodes.

The image is made up of the IMAGE SUBTREE, the TRANSFER FUNCTION, which says where to connect the subordinates of the occurrences of the nodes which do not themselves appear in the schema, and the ASSIGNMENT PART, which computes the values of the variables appearing on the image nodes.

The example shown in Figure 10 is a rule to handle negation and is used at the beginning of the analysis. If the morphological analysis has produced a compound word, the node corresponding to the 'NE-' will be erased and the negative information transferred into the variable NEG on the principal node. A1E is the name of the rule.

The schema has three nodes, arbitrarily called 1, 2 and 3 (any identifier could be used). Its shape is shown on the left. The star '*' indicates that 2

AIE: (*1, &niv=3)	}	Schema
1(2, *, 3) /		
1: \$ULMCP; 2: UL -E- 'NE-' /	}	Image
==1(3) / *←2 /		
3: 3, NG:=NE.		

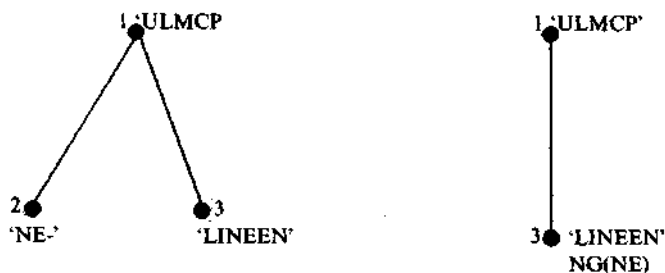


Figure 10.

and 3 must be contiguous in the object tree. In the expression ‘(*1, &niv=3)’, ‘&niv=3’ says that the root of the schema must be on level 3 in the object tree (the root being at level 0) and ‘*1’ says that node 1 is invariant in the transformation, so that it may be shared by another rule. This feature is called HORIZONTAL PARALLELISM. For example, on the subtree corresponding to ‘NELINEJNO-NEPRERYIVNO’, there are two occurrences of AIE, and they may be executed in parallel.

More generally, the root of the transformation, ρ_T , may be any node in the schema. The subtree rooted by ρ_T is the ‘active part’ while the rest is called the ‘context’. Several occurrences of rules may share contextual nodes. Rules using this facility are analogous to context sensitive rules for strings.

‘1:\$ULMCP’ asks if the UL of node 1 is equal to ‘ULMCP’ by calling the procedure ULMCP (confusion is avoided by the \$ prefix), which has been declared as ‘ULMCP==UL -E- “ULMCP”’. ‘-E-’ is the equality operator. In the image part, ‘*←2’ says that possible subordinates of node 2 will be lost (transferred to the ‘empty’ node), and ‘3:3, NEG:=NE’ says that the variables on the new node ‘3’ will be computed by taking the ones of the old node ‘3’ and then assigning the value NE to the variable NEG.

2.1.3. *Transformational grammars (TG) and control graph.* The TRs form an unstructured set. A TG is an ordered set of TRs. This order is local to each grammar, and is mainly used to resolve conflicts between rules.²

The control graph is a loop-free oriented and ordered (multi) graph where each node bears the name of a TG, with its modes of application, or the exit symbol '&nul'. The arcs of the graph may bear conditions which must be met in order for the arc to be traversed.

Execution of a transformational system (TS) occurs as follows. An initial node is chosen in the graph. Using a built-in back-tracking algorithm, the automaton traverses the control graph through the first possible path leading to an exit symbol, whereby each traversed TG is applied to the object tree. When the application of a TG is completed, the leaving arcs whose conditions are not met by the resulting object tree may not be traversed. The control graph of the TS written for the analysis of Russian is given in Figure 11 below. Note that, if there is no such possible path, the output will be set equal to the input.

Rules are usually grouped in grammars according to the following criteria:

- they correspond to related linguistic phenomena.
- they express transformations used for a certain logical step of the linguistic process (multilevel analysis in this section).
- they share the same execution modes (for instance rules which may iterate will appear in exhaustive grammars, and rules which don't will rather be put in unitary grammars). This criterion is essentially strategical.

This architecture makes it possible to:

- limit the interaction between rules and avoid many combinatorial problems
- develop strategies and heuristics
- test and modify the TGs separately (different control and tracing parameters may be associated with each TG).

Some of the execution modes of a TG express its behaviour with respect to the traversal of the control graph, some control iteration, and the rest determine the choice between possible occurrences.

If a TG is in 'imperative' mode ('I') and no transformation has occurred, the path is blocked and the automaton must back-track. If a TG is in 'zero' mode ('Z'), the automaton may not back-track before this TG, so that previous work on the object tree may not be undone.

2.1.4. Application of a TG. Elementary application. An elementary application of grammar $G: R_1, R_2, \dots, R_p$ occurs conceptually as follows (we ignore short-cuts used for some combinations of execution modes). The automaton computes the family O_j of all possible occurrences of the R_i s in the object tree. Let P_j be the nodes corresponding to the roots of the associated transformations. Then a maximal subfamily of the O_j s is selected so that no active node is shared by two occurrences, and in accordance with the 2 following execution modes:

— in mode 'B' (bottom), preference is given to the deepest P_j s, as opposed to mode 'H' (high).

— in mode 'C' (cut), no P_j may strictly dominate another P_k , whereas occurrences may be 'embedded' in 'T' (total) mode.

At that point, some nodes of the object tree may still be roots of several transformations. The conflicts are solved by using first the priorities induced by the order R_1, \dots, R_p , and then, for each rule, the constraints given by the associated schema. Several occurrences (possibly of different rules) may still be rooted on the same node, if the corresponding P_j s are 'contextual' (invariant in the transformation, like node 1 in Figure 10).

Then, the corresponding rules are applied in parallel on the object tree. If P_j is the root of an occurrence of R_k , the automaton marks R_k on P_j , and, in 'D' (down) mode, the rules of higher priority (R_1, \dots, R_{k-1}). If R_k gives rise to a recursive call, it is performed on the subtree rooted by P_j .

Iteration. If a TG is in 'E' (exhaustive) mode, elementary applications are iterated until no rule is applicable. In 'U' (unitary) mode, only one application is performed. In 'G' (guarded) mode, as opposed to 'L' (liberated) mode, if R_k is marked on a node, R_k may not be used on and below this node, so that the number of potential applications diminishes.

Recursion. In G , R_k may recursively call the SUBGRAMMAR of G' consisting of rules R'_{k1}, \dots, R'_{kr} of G' , and inheriting its execution modes. We write then: ' $G: R_1, \dots, R_k (G'/R'_{k1}, \dots, R'_{kr}/ \text{validated nodes}), \dots R_p$ '. If all rules of G' appear, we write: ' $\dots R_k(G'/*/\text{v.n.}), \dots$ '. Validated nodes will be explained below.

R_k may also recursively call the transformational subsystem with initial node G' , in which case we write: ' $G: R_1, \dots, R_k(*G'/*/\text{v.n.}), \dots R_p$ '. In G , each rule may have one (and only one) recursive call.

In order to execute a recursive call, the system cuts out the subtree dominated by P_j and handles it as a new object tree. The image nodes which are not validated may not participate actively in any transformation during the call. This reduces the number of potential occurrences and enforces decidability. The result of the call replaces the input subtree in the original tree.

Recursive calls, especially of the second type, make it possible to use different strategies on different parts of the unit of translation. However, the second type, having been introduced in 1979, has not been used in the RUS-FRA application.

Modes of execution. We give them all, for the sake of completeness. They are all binary, the italicized value is by default, and any combination is valid.

Unitary	Bottom	Cut	Down	Guarded	Facultative	Normal
Exhaustive	High	Total	Punctual	Liberated	Imperative	Zero

2.2. Multilevel analysis of Russian

2.2.1. Overall organization. Figure 11 shows the control graph of the TS.

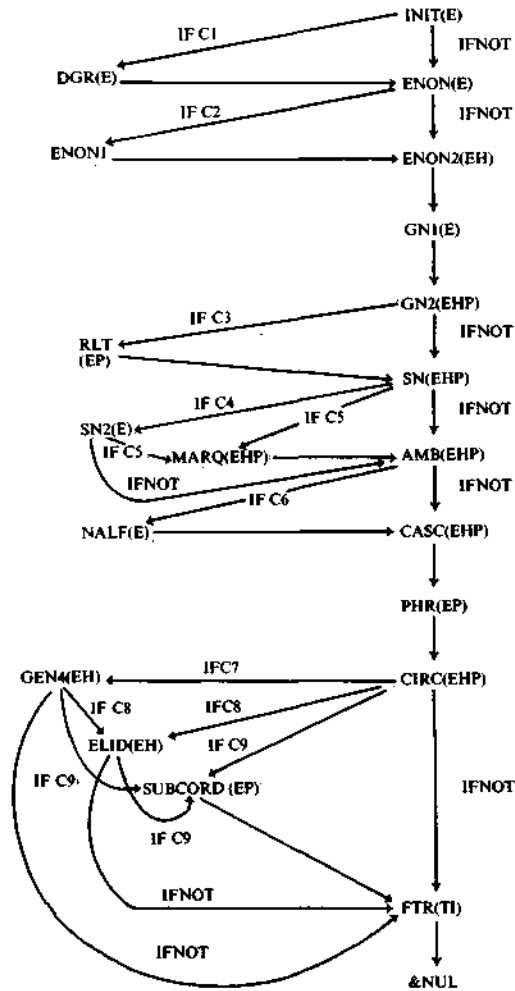


Figure 11.

Content of conditions:

- C1 — degree or demonstrative elements in the text.
- C2 — comma or semicolon.
- C3 — Relative or participial clause.
- C4 — Preposition or long-form adjective outside the clause.

- C5 — Infinitive or subordinate clause.
- C6 — Non-alphabetical form.
- C7 — Genitive nominal group outside the clause.
- C8 — Isolated long-form adjective.
- C9 — Subordinate clause.

2.2.2. *Grammars.* Let us now explain the functions of each grammar.

INIT Its aim is to homogenize the input and to simplify it. More specifically:

a. One suppresses certain parasitic information owing to morphological analysis particularities:

— value of animate accusative for inanimate nouns, e.g.

STOLA

— value of inanimate accusative for animate nouns, e.g.

DRUZQYA

— certain propositional government code in passive or reflexive verb forms, e.g.

NABLYUDATQ ZA

b. The subtrees resulting from prefixation or compound words are transformed by attaching the secondary elements under the main word; in case of the prefix NE the node corresponding to it is suppressed and the negation is expressed by a special variable.

c. One treats as compound words the non-alphabetical numbers, e.g. 543, where the last figure is considered as the main word, and the others as prefixes.

d. The initials of first names are treated as prefixes in relation to the family name.

e. One proceeds to identify all clauses in parentheses and the cases of simple noun enumeration, e.g.

NEJTRON, PROTON I DEJTRON

in order not to consider the commas as clause limits.

DGR This grammar is called only if a corresponding configuration is present in the text. Its aim is to recognize an analytic expression of degree, e.g.

BOLEE PONYATEN

SAMYIJ TRUDNYIJ

TAKOJ ZHE

and to replace the secondary nodes by marking the degree values on the principal node of the clause. The same grammar allows us to search for expressions such as:

VSE BOLEE I BOLEE
1 2 3 4

which is considered as an adverbial modifier case and not as a comparative form. The resulting structure keeps only the nodes No. 1 and 2. We also solve the ambiguity of KHEM after a comparative word, and of DRUG before DRUG preceded or not by a preposition.

ENON-ENON1-ENON2 This set of grammars aims at breaking down each sentence into so called 'utterances' corresponding to linguistic units which are textually marked. More specifically, we use typographical markers like commas, conjunctions etc., as delimiters for these utterances. For example, this sentence:

VIDIM DOM, GDE ON ZHIVET

will be broken down into two utterances:

1. VIDIM DOM,
2. GDE ON ZHIVET.

The node dominating each utterance is marked by '*ENONCE', which is a special lexical unit. Moreover, the ENON2 grammar recognizes any relative or participial clause corresponding to the utterances already found and marks them with special variable values.

GN1 In this grammar, we try to build the simple noun group from either a long-form adjective or a numeral or a preposition before a substantive form. At the same time, we recognize any adverbial expression formed by the word OBRAZOM after an adjectival form. E.g.

SLEDUYUTHIM OBRAZOM

The result of this construction corresponds to the notion of a pure noun group or a prepositional group respectively marked *GN or *CIRC as special lexical units and GN or GP as syntactical classes. The same GN1 grammar also solves certain nominal ambiguities between genitive singular and nominative plural. E.g.

POPYITKA ISSLEDOVANIYA

GN2 The GN2 grammar continues to search for other elements dependent on the noun group:
— any secondary element subordinated to the adjective absorbed by a noun group, e.g.

NEKOTORYIX PROVEDENNYIX OPYITOV
XOROSHO GN
PREDUSMOTRENNYIX IMI

— any secondary element subordinated to the noun group, e.g.

IZ PROVEDENNYIX OPYITOV
PYATQ PROVEDENNYIX OPYITOV

In other respects, the GN2 grammar fixes the correct interpretation of certain ambiguities like, e.g.

VOLNA DLINOJ

where the instrumental case marks the measure, or like, e.g.

NI S KEM
KOE S KEM

where the interrogative elements become negative or indefinite, or like, e.g.

POSLE

which loses its nominal interpretation as AMBASSADOR and preserves only its prepositional one as AFTER. Moreover, this grammar studies the expressions like:

PERVYIJ IZ AVTOROV

where the first word represents an elliptic use of the plural noun after IZ and may be interpreted as:

PERVYIJ AVTOR IZ AVTOROV

Lastly, every isolated and nonambiguous noun is transformed into a normal noun group.

RLT The aim of this grammar consists of searching for the nominal antecedents of relative and participial clauses constructed by the ENON2 grammar. In the case of relative pronouns, their lexical unit is replaced by the corresponding antecedent value, while the participial clause takes in a supplementary node with the same lexical unit as its antecedent. E.g.

DOM, KOTORYIJ STROITSYA becomes
DOM, DOM STROITSYA and
DOM, POSTROENNYIJ NAMI becomes
DOM, DOM POSTROENNYIJ NAMI.

SN The SN grammar searches for the personal verb as the main element of the utterance, in order to copy onto the *ENONCE node all

information carried by the verb. At the same time, it finds modifying verbal elements like negative or conditional particles and auxiliary verbs. E.g.

DOLZHEN BYITQ POLUKHEN
NE BUDET IGRATQ
SKAZAL BYI.

SN2 In this stage of our analysis, we perform two tasks. The first consists in solving ambiguities between the short-form neutral adjectives and the adverb, where it is possible. E.g.

YETA KNIGA *XOROSHO* OPISYIVAET

The second task consists of building the compound noun group from the groups already built. We call it noun group embedding. E.g.

IZ POLUKHENNYIX RANEE AVTORAMI REZULQTATOV

On the other hand, the adverbs which were ambiguous with short-form adjectives may be tested for prepositional group constructions like:

SOGLASNO POLUKHENNYIM REZULQTATAM

MARQ This grammar is designed to build all types of subordinate verbal and infinitive clauses. If possible, the ambiguity between the adverb and the short-form adjective is solved: e.g.

ESTESTVENNO, KHTO or
TRUDNO PONYATQ

In the case of an infinitive clause, we create a special utterance, just as for a personal verb clause.

AMB The AMB grammar searches for the most important terms of the clause, namely subject, object and near dative complement. We begin with the non-ambiguous noun groups, like:

TAKOE POLOZHENIE OPISYIVAET *KNIGA*.

After finding the subject *KNIGA* we can say that the first ambiguous noun is the object. On the other hand, we continue to solve the ambiguity between adjectives and adverbs, using the newly constructed verbal utterances. The same applies to constructions like:

ONI SOOTVETSTVUYUT FUNKCIJ
subject dative

NALE Here, the non-alphabetical forms are treated as appositions or verbal complements.

FUNKCIYA f(x)

CASC In this grammar, we examine all cases of genitive nominal embeddings, like:

TEORIYA ISKHEZNOVENIYA DAVLENIYA

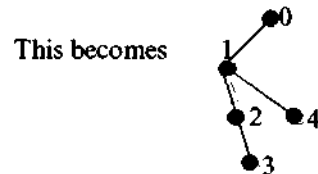
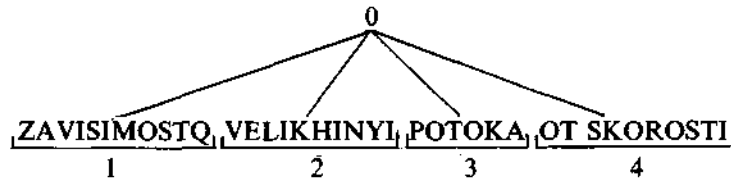
The resulting structure is provisional, because genitive nouns are attached to the non-ambiguous noun group.

PHR The PHR grammar marks all strongly-governed groups subordinated to the utterance, like agent, patient, attribute, etc. E.g.

YETO	OBWYASNYAETSYA	AVTOROM
Subject		Object
Patient		Agent

Wherever possible, the dependent groups are labelled with corresponding markers with regard to the syntactical class and the logical relation.

CIRC-GEN4 Both these grammars realize the distribution of prepositional and genitive noun groups between their noun heads, according to several semantic and syntactic criteria. Up to the present, all these groups are represented in the provisional structure, on the same level: e.g.



ELID Here, one searches for the possible antecedent of pronominal expressions as well as isolated adjectives; e.g.

VZYATQ VTOROJ BLOK, PREZHDE KHEM ISPRAVITQ
ISPORKHENNYIJ

We generate a noun group by copying the lexical unit of the antecedent. If the elliptic element is not a personal pronoun, it becomes qualifier or

determiner according to its syntactic class. The new noun group is tested so as to find its semantic and logical relation within the utterance.

SUBCORD The existence of this grammar is purely formal. It serves to modify the hierarchy of different subordinate clauses. Indeed, all of them are represented on the same level as the principal sentence. E.g.

KNIGA₁, KOTORUYU MYI KHITAEM₂, INTERESNA₃

Sentence No. 2 descends under sentence No. 1, which in its turn is attached to sentence No. 3. After this, we test the relations between the elements of sentence 1 and the verb of sentence 3.

FTR This last grammar copies calculation results from each non-terminal node onto its terminal governor. On the other hand, the logical relation computed for each prepositional group is copied onto the terminal preposition node (head). These operations prepare the lexical transfer step, where context is limited to one node at a time.

2.2.3. *Examples.* Let us now show the construction of a simple prepositional group. The rules correspond to the context-free rules:

GN1 → ADJ + NOUN

GN2 → ADJ + GN1

GP → PREP + GN_i (i=1,2) (see Figure 12)

Nodes A and B correspond to contiguous occurrences. Node 1 must be an adjective in the long form (procedure AQFOL) and 2 must be a noun. Procedure CGN, called on 1 and 2, asks for agreement in case, gender and number. In the result, possible ambiguities are lost (*←A, B), node B takes the variables of the noun, and values are determined by the call to AFCGNA on the (old) nodes 1 and 2. Node 2 is assigned format F (constant mask of variables), then altered by the call to the procedure AFGOV. RESEP is a conditional procedure which assigns values to new node 1 from values of variables on old node A: here, it assigns the auxiliary UL '*EPIT' to A. This 'lexicalizing' of the syntactic function takes advantage of the fact that non terminal nodes have no 'natural' lexical unit, and helps debugging, because the value of the UL is shown on the graphic displays of trees (in all components of the system), as shown in Figure 13.

Note that the node 'TERMO', subordinate of node 1, has been transferred under node 1 in the image. Note also that several transformations of that kind may occur in a parallel way in an utterance (*énoncé*), because node 0 is contextual (*0).

AQN1: (*0, &niv=2)
 0(A(1), *, B(2))/
 0: \$ENØNCE; A: \$ULØCC; B: \$ULØCC;
 1: \$AQFØL; 2: \$NM/\$CGN (1, 2)
 ==0(B(A(1), 2))/*←A, B; A←*; B←*/
 B: 2, \$AFGNA (1, 2); A: 1;
 1: *F, \$AFGOV (1);
 2: *F, \$AFGOV (2), \$\$RESEP (A; 1).

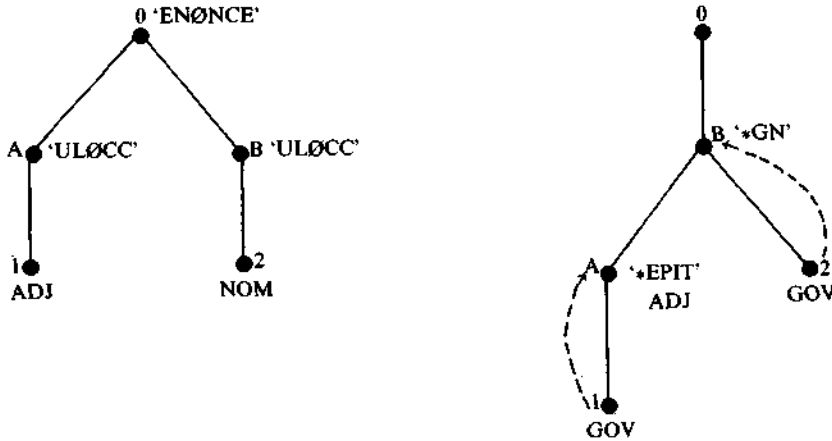


Figure 12.

Note also that the ambiguity between *SISTEMYI* as genitive singular (33) and nominative plural (34) has been correctly resolved.

In rule AQN2, the transfer expression '2←(OD) 2' says that subordinates of old node 2 (dotted triangle) will be connected to the right of the subordinates of (new) node 2 in the image (see Figure 14). By default, they would have been connected to the left of them. This rule applies on nodes 21(27(28),29), yielding the subtree in Figure 15.

Finally, rule GP2 (Figure 16) constructs a prepositional group, appearing in the subtree in Figure 17.

3. Lexical transfer

This step is the only one in which two sets of variables, one for the source language and one for the target language, are accessible. The second set includes variables taken from the first (pivot variables and lower level source variables used for contrastive processing in the following step), as well as new variables, relative to the target.

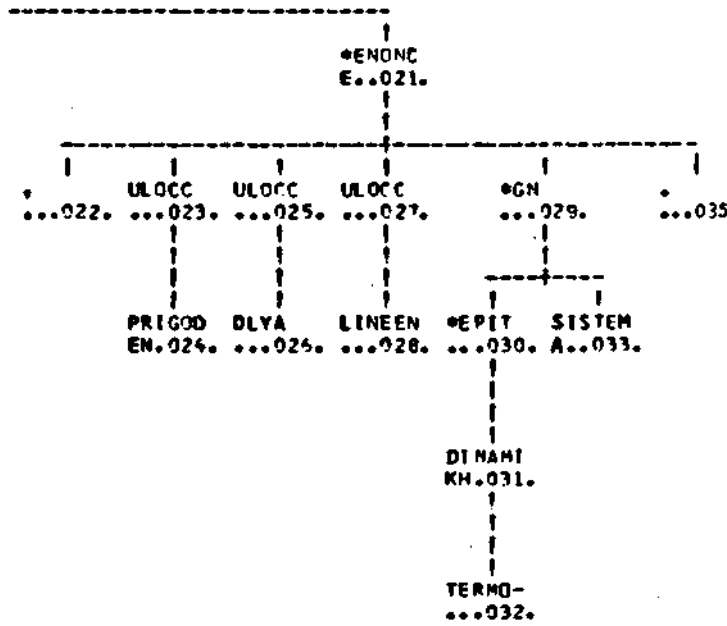
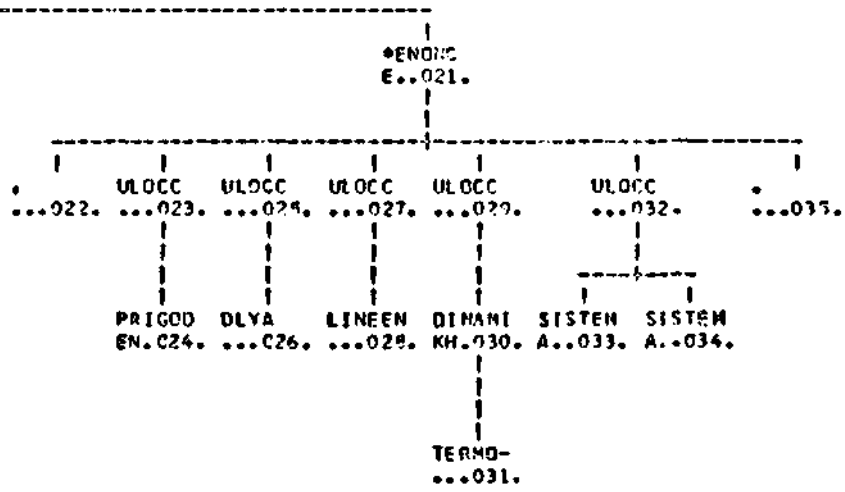


Figure 13.

```

AQN2: (*0, 0=&NIV<-2)
      0(A(1), *, 2)/
      A: $ULØCC; 1: $AQFØL;
      2: $GN/$CGN(1, 2)
      ==0(2(A(1)))/

      *←A; A←*; 2←(ØD)2/
      A: 1;
      1: *F, $AFGØV(1);
      2: 2, $AFCGN(1, 2), $$RESEP(A; 1).
    
```

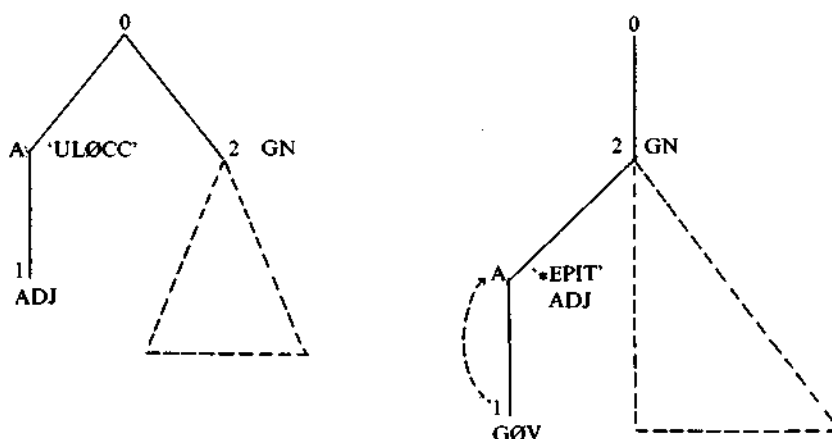


Figure 14.

3.1. The algorithmic component: TRANSF

Algorithmically speaking, the model is essentially a bilingual dictionary of simple transfer rules accessed by the source UL. Each rule is a sequence of 3-tuples (condition, image subtree, assignments), the last condition being empty (NIL).

The automaton traverses the input tree in preorder, creating the object tree as follows. The UL of the current node is used to access the dictionary. The first triplet of the item whose condition is verified is chosen. The image subtree (generally consisting of only one node) is added to the output, with values of variables computed by the assignment part.

Hence, the output tree is very similar to the input tree. The possibility of transforming one input node into an output subtree may be used to create compound words or to create auxiliary nodes used in the following step (structural transfer) to treat idioms.

As this model is algorithmically very simple, it is the only one where no

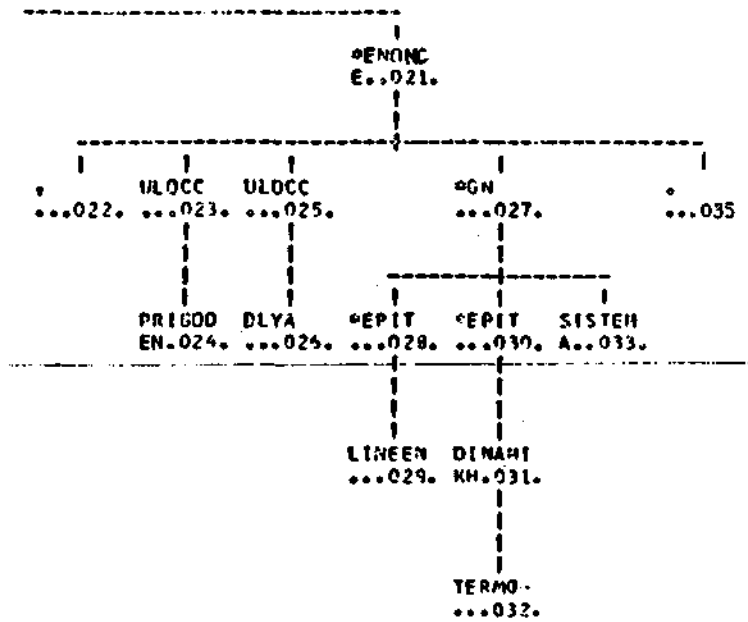


Figure 15.

GP2: (*0, 2=&NIV)
 0(A(1), *, 2)/
 0: SENØNCE; A: SULØCC;
 2: SGN; 1: SPP/\$C (1, 2)
 =-0(2(1))/+←A; 2←(ØD)2/
 2: 2, SAFGP (1, 2);
 1: 1, SAFPP (1, 2).

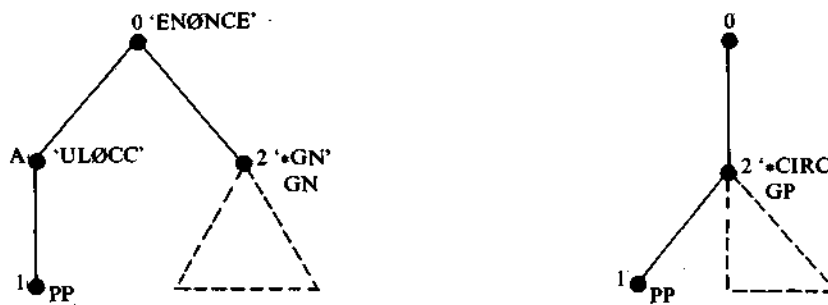


Figure 16.

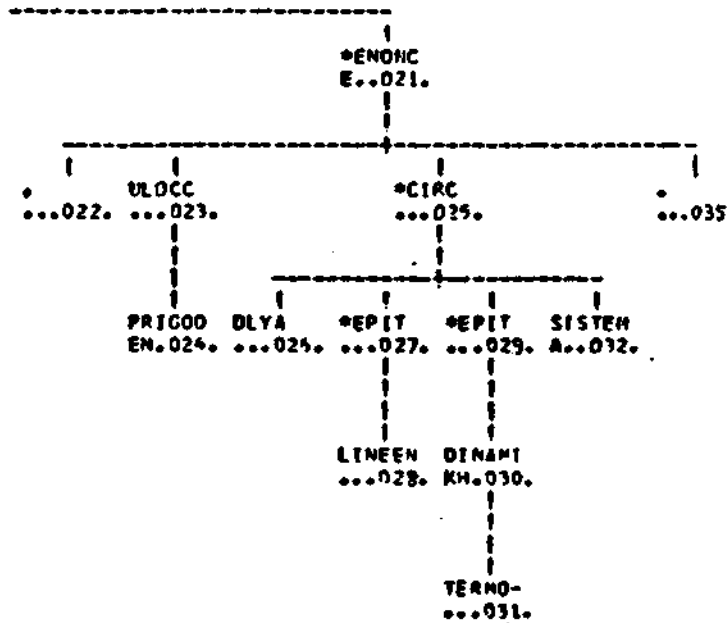


Figure 17.

trace is provided. Figure 18 gives an idea of the metalanguage of the dictionary.

'0(1,2)' describes the image subtree for 'NAPRIMER'. The other ones are reduced to one node (default). '+VBF1' says that the non-null values of variables in format VBF1 will be copied into the target node. RFPF is an assignment procedure. '*PP' says that all variables of format PP (except the UL) will be copied onto node 1.

```

'FORMULIROVATQ' == / /'FORMULER' ,+VBF1,$RFPF.
'PRICIP' == / /'PRINCIPE' ,SNMAS.
'PRIPOMOTHI' == / /'A-L-AIDE' ,+MPCD.
'NAPRIMER' == /0(1,2)/0: '*LOCF' ,+VIDE;
1: 'PAR' ,*PP;
2: 'EXEMPLE' ,*NMMS.
    
```

Figure 18.

3.2. Lexical transfer from Russian to French in *RUS-FRA*

3.2.1. How to choose equivalents. There are two situations:

- a. to one source UL corresponds only one target UL. No question is asked, and the item is quite simple:

'KABINA' = // 'CABINE'.

b. to one source UL correspond different target ULs and we can choose by asking some questions on the source node, as in:

'KAKOJ' = = \$P2A // 'AUCUN' /
// 'QUEL'.

(P2A is true if the input node is negative).

These questions are expressed as conditions on the (source) variables present at the end of the multilevel analysis.

3.2.2. *Representation of the equivalents.* In most cases, the image subtree is degenerate (reduced to one node). More complex subtrees are used in the following eight situations.

a. translation of fixed idioms, which have been assigned one UL in the source:

'KAK-I' = = /0(1,2)/ 0: '*LOCF'; 1: 'DE-MEME'; 2: 'QUE'.

'POYETOMU' = = /0(1,2)/ 0: '*LOCF'; 1: 'C-EST'; 2: 'POURQUOI'.

b. separation of target prefixes, in order to reduce the number of the target ULs:

'OKOLOZVUKOVOJ' = = /0(1)/ 0: 'SONIQUE'; 1: 'PRE-'.

c. creation of compound words corresponding to simple source lexical units:

'ZVENQEVOJ' = = /0(1)/ 0: 'CHEF'; 1: 'GROUPE'.

d. creation of compound predicates:

'PREPYATSTVOVATQ' = = /0(1)/ 0: 'OBSTACLE'; 1: 'FAIRE'.

e. verbal government — see 3.2.3.

f. unsolved polysemy — see 3.2.4.

g. non fixed idioms — see 3.2.5.

3.2.3. *Verbal government.* In many cases, the French verbal government may be computed from the Russian one by using a simple table. For instance, dative→A and NA→SUR. In case of an exception, we use a supplementary node containing the UL of the French preposition and a trace of the Russian government, for subsequent use in the structural transfer:

'ZAMENITQ' = = /0(1)/ 0: 'REPLACER'; 1: 'PAR', PG=2NA.

2NA indicates that PAR corresponds here to NA+accusative.

3.2.4. *Polysemy.* There are of course cases where we are unable to choose the correct equivalent by using only the information computed by the analysis. In some cases, a more sophisticated third-generation system might use adequate world-knowledge to resolve the ambiguity. In other cases, the ambiguity would persist, and perhaps also even for human translators. When in doubt, we choose to preserve the ambiguity and to generate the two (or more) possibilities in the output text, with an adequate presentation (using parentheses, or slashes ...).

'ZAPUSTITQ' = = /1(2)/ 1: 'LANCER'; 2: 'NEGLIGER', \$HOM.

Procedure HOM assigns a special value to a (tactical) variable on node 2, whose interpretation is that 2 is a synonym and should be generated in a particular way.

3.2.5. *Source idioms.* Non-fixed source idioms appear in the result of the analysis as subtrees. As context in TRANSF is reduced to one node, they cannot be correctly translated at this stage of the process. The method here is to make PREDICTIONS for the case where idioms are indeed present, and to confirm or infirm these predictions during the structural transfer. For example:

1. 'KOYEFFICIENT' = = /0(A)/ 0: 'COEFFICIENT';
A: '+KOYEFFICIENT'.
2. 'POLEZEN' = = /0(X(A(1)))/ 0: 'UTILE'; X: '*RLX3';
A: '+KOYEFFICIENT'; 1: 'RENDRE'.
3. 'DEJSTVOVATQ' = = /0(X(A))/ 0: 'AGIR'; X: '*RLX3';
A: '+KOYEFFICIENT'.

for: KOYEFFICIENT POLEZNOGO DEJSTVIYA → RENDEMENT.
(1) gives the normal translation, plus the information that this lexical unit may be the reference of an idiom; (2) gives also the type of rule to be used in structural transfer; and (3) has been chosen as KEY (or HEAD) of the idiom. It contains the translation of the idiom, to be produced if and only if the other components are there and the conditions (expressed in rule RLX3) are met. If not, nodes X,A,1 will be suppressed and the normal translation produced (that is, only node 0 will be retained).

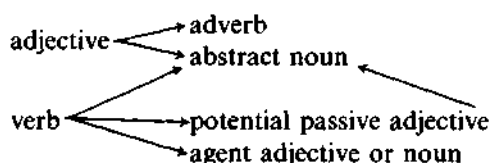
3.3. *Nature of the information given by the dictionary*

The information, relative to the target language, must be sufficient for the steps of structural transfer and syntactic generation, as no dictionary is accessed during these steps. This information may be general or relative to a particular morphological class.

3.3.1. *General information.* It consists of:

- a. an associated morphological class (noun, adjective, verb or other), chosen in this step because there are different types of flexional systems.
- b. derivational codes, indicating the possible derivations from the 'main' morphological class associated with the lexical unit, in terms of productive suffixes.

The possible derivations are:



The abstract noun may be masculine or feminine; for each case, we note the appropriate suffix:

for feminine: -ION, -URE, -ERIE, -ANCE or -ENCE, -ESSE, -EUR, -ITE or -ETE or -TE, -E (from a basic verb), and the coincidence with the participle (e.g.: mêlée).

for masculine: -EMENT, -AGE, infinitive of basic verb, its participle, -ISME.

The agent derivation may be of three types: -EUR, -RICE after T; -EUR, -EUSE or irregular on -IER; and -ANT or -ENT.

- c. negation code for: IN-, IM-, IL-, IR- prefixes; NON- prefix; and an expression with PEU.

3.3.2. *Information proper to a particular morphological class.*

Noun. We give:

- the gender
- the number, if constant
- the preposition to use in locative expressions (*en France, au Japon, à Madagascar*)
- the fact that the noun is directly derived from the adjective (substantivized): 'FIZIKA' = // 'PHYSIQUE'.

Place of adjective or adverb. Adjectives which must occur before the noun are noted, as well as their behaviour relative to determination (MON replaces the article, TEL wants an indefinite article, and ORD and BEL note ordinals and ordinary qualifiers).

Adverbs are divided into three classes, with prototypes TRES (always before the adjective or the qualifying adverb), TOUJOURS (inserted between the auxiliary and the participle in compound tenses), and TOT (remains at the same place as in Russian).

Verb. a. ACTUALIZATION MARKS concern:

— the simple tenses: present (indicative or subjunctive), imperfect (indicative), simple future, preterite (indicative or subjunctive).

— the modes: indicative, subjunctive, conditional, imperative, participle, gerundive, infinitive.

— the degree of tense composition: 0, 1 (*a parlé*) or 2 (*a eu parlé*).

b. SYNTACTIC REQUIREMENTS are:

— subjunctive after the verb or after a conjunction

— conditional

— preposition A before subordinated infinitive (*donner à faire*)

— no preposition before subordinated infinitive (*vouloir faire*)

— indirect complement in Russian becomes direct in French

— Russian instrumental becomes DE+GN

— NA+Russian accusative becomes DE+GN

— K+Russian dative becomes DE+GN.

c. NECESSARY CHANGES IN THE ARGUMENTS while translating are of 2 kinds:

— Russian impersonal becomes personal: MNE XOKHETSYA → JE VEUX.

— Russian agent becomes object: MOZHNO DELATQ → ON PEUT FAIRE.

d. AUXILIARY AND VOICE are given:

— ETRE or AVOIR

— Russian reflexive becomes passive, active, or reflexive in French.

Construction of the governor³ (preposition or conjunction). It may be simple or composed with DE, QUE or A: *afin de, afin que, de sorte à, de sorte que.*

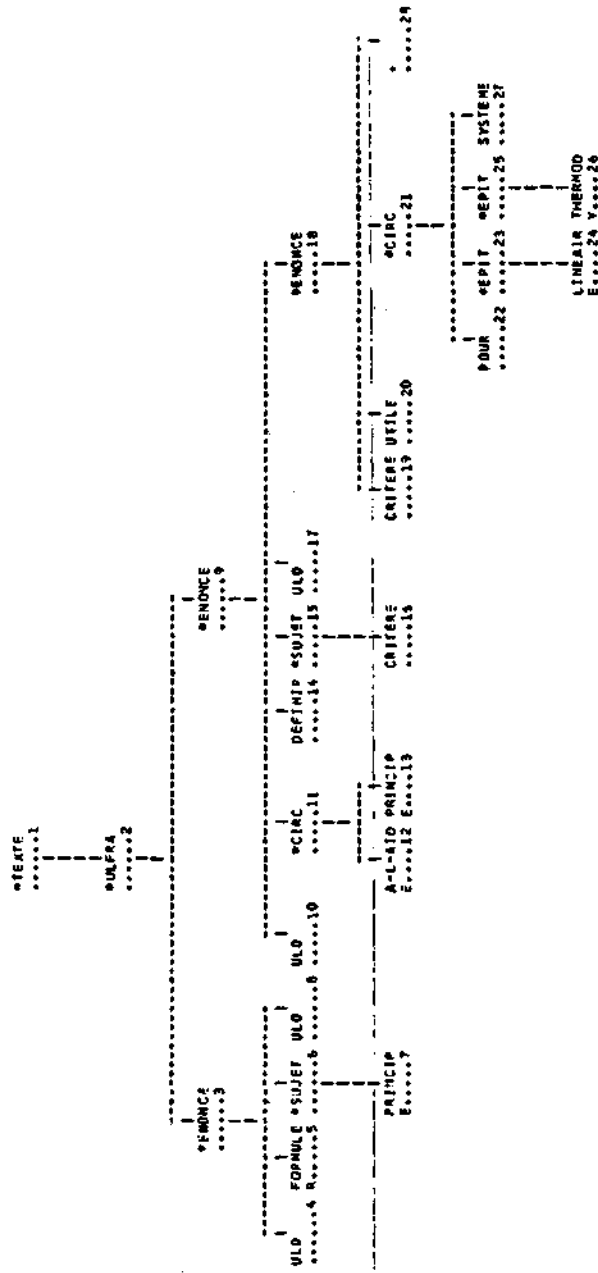
Pronoun. It may be an atonic subject (JE, IL, ON), an atonic complement (ME, LE), or a tonic complement (EUX).

Construction of a prepositional group equivalent to one Russian word. Such groups use DE, A or EN (TRANSPORTNYIJ → DE TRANSPORT, GAZOVYIJ → AU GAZ).

Preposition in locative expression. It may be EN, A, DANS (*en France, à Madrid*).

3.4. Example

Figure 19 is the result of this step on our standard example.



```

SUMMET 1 : UL(*TEXTE).
SUMMET 2 : UL(*ULFRA).
SUMMET 3 : *SPFORMULTROVAN: UL(*ENDONCE),KI(PH),K(AQ),A(PI),G(M),N(S),P(1,2,3),R(F(PF),ABS(A2,SJ),T(PAS),MD(PRT),
CP1(ACC)).
SUMMET 4 : UL(ULO),FT(D68).
SUMMET 5 : UL(*FORMULER),L(XGOV),K(AQ),K(F(VB),SXFI(ON),R(LR(F3),A(P),G(M),N(S),P(1,2,3),R(F(PF),ABS(A2,SJ),T(PAS),
MD(PRT),CP1(ACC)).
SUMMET 6 : PRINCEP: UL(*SUJET),K(IGM),NAT(OBJ),AG(A2),K(NM),S(M),N(S),P(3),PG(60B),ANM(INA).
SUMMET 7 : UL(*PRINCIPE),NAT(OBJ),AG(A2),L(XGOV),K(NM),K(F(NM),G(M),N(S),P(3),ANM(INA).
SUMMET 8 : UL(ULO),FT(FIN).
SUMMET 9 : OPR(*DLYAETSYA: UL(*ENDONCE),KI(PH),AG(A6),K(VB),A(I),G(F,M,N),N(S),P(3),R(F(R),PG(4IZ),ABS(A2,SJ),T(PRE),
MD(VRB),CP1(ACC),CP2(PRP).
SUMMET 10 : UL(ULO),FT(D68).
SUMMET 11 : UL(*CTAG),K(IGP),ANF(ULT),K(NM),G(M),N(S).
SUMMET 12 : UL(*L-AIDE),C(SACT),FT(PP),K(PP),ANM(INA),MDC(OE).
SUMMET 13 : UL(*INDIPE),ANF(ULT),L(XGOV),K(NM),K(F(NM),G(M),N(S),P(3).
SUMMET 14 : UL(*DEFINTR),AG(A6),L(XGOV),K(VB),K(F(VB),SXFI(ON),A(I),G(F,M,N),N(S),P(3),R(F(R),ABS(A2,SJ),T(PRE),
MD(VRB),CP1(ACC)).
SUMMET 15 : KRITERIJ: UL(*SUJET),K(IGN),C(SACT),AG(A2),K(NM),G(M),N(S),P(3),ANM(INA),CP1(DAT).
SUMMET 16 : UL(*ITERE),C(SACT),AG(A2),L(XGOV),K(NM),K(F(NM),G(M),N(S),P(3),ANM(INA),CP1(DAT).
SUMMET 17 : UL(ULO),FT(FIN).
SUMMET 18 : PRIGOPNYI: UL(*ENDONCE),K(MO),C(SACT),AG(A6),K(AQ),G(M),N(S),PG(5<),ANM(INA),CP1(PRP).
SUMMET 19 : UL(*ITERE),FT(D68),ANF(ULT),K(NM),K(F(NM),G(M),N(S).
SUMMET 20 : UL(*UTILE),C(SACT),AG(AQ),L(XGOV),K(AQ),K(F(AQ),K(F(ITE),PRG(AJQ),G(M),N(S),G(M),N(S),ANM(INA),MD(FIN).
SUMMET 21 : SISTEMVI: UL(*CIRC),K(IGP),NAT(OBJ),K(NM),G(F),N(S),P(3),PG(5PO),PG(4DLA),ANM(INA).
SUMMET 22 : OLYA: UL(*DUR),NAT(OBJ),FT(PP),K(PP),K(F(V),ANM(INA).
SUMMET 23 : MELINEJN: UL(*PITI),K(IG),C(SOAL),K(AQ),G(F),N(S),NG(NE),MDIADJ).
SUMMET 24 : UL(*INDIPE),C(SOAL),L(XGOV),K(AQ),K(F(AQ),SXFI(ITE),PRG(AJQ),G(F),N(S),NG(NE),MDIADJ).
SUMMET 25 : TER-MODINAMIKESKOJ: UL(*PITI),K(IG),C(SALT),K(AQ),G(F),N(S),MDIADJ).
SUMMET 26 : UL(*THERMODYNAMIQUE),C(SOAL),L(XGOV),K(AQ),K(F(AQ),PRG(AJQ),G(F),N(S),MDIADJ).
SUMMET 27 : UL(*SYSTEME),NAT(OBJ),L(XGOV),K(NM),K(F(NM),G(M),N(S),P(3),PG(4D_A1),ANM(INA).
SUMMET 28 : UL(*),FT(FIN),K(V6).

```

Figure 19.

4. Structural transfer

The algorithmic component used in this step is ROBRA, which has been presented in 2. The aim of this step is to realize all transformations of contrastive nature, so as to produce the desired intermediate target structure as output.

4.1. Structural transfer from Russian to French

4.1.1. Overall organization. Figure 20 shows the control graph of the TS (transformational system).

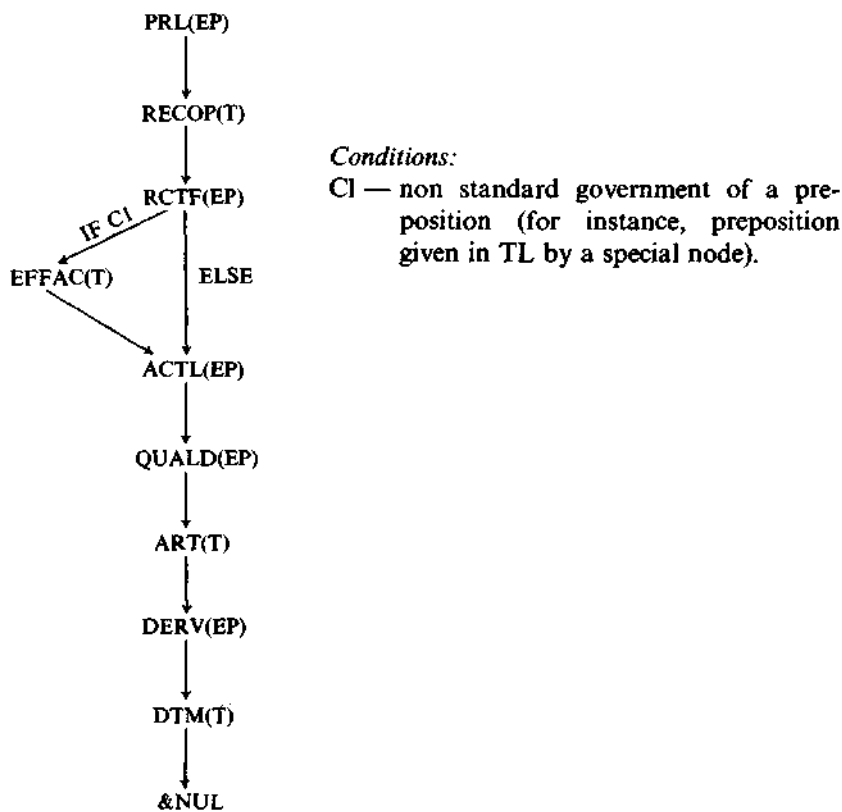


Figure 20.

4.1.2. Grammars. Let us now briefly explain the function of each transformational grammar.

PRL — In case of failure of the analysis, certain 'ULOCC' nodes may persist. Appropriate values of variables are assigned to their dependents (unanalysed words).

— Idioms (predicted in TL) are handled: the context is verified and the corresponding idiom is either conserved or discarded.

— the determination of the nominal groups is computed, in order to prepare the generation of articles.

RECOP — Certain information is copied from terminal 'head' nodes (*gouverneurs*) on the non terminal representing the group: required mode, type of adjective, postponed preposition, inversion of arguments.

Note: By analogy with compilation theory, one might say that grammars PRL and RECOP compute 'synthesized attributes' (bottom-up in the tree).

RCTF — In case of non-standard government, the context is checked, and the preposition given (in TL) under the headnode of the group replaces the standard one:

● fabriquer + de → fabriquer + en.
|
● en

— Information relative to the translation by DE of certain complements is copied onto the non-terminal node. For instance:

POKRYITQ + instrumental → COUVRIR DE
UVELIKHITQ NA + accusative → AUGMENTER DE
BLIZKIJ K + dative → PROCHE DE

A similar case occurs for the translation of an indirect object by a direct one:

POLQZOVATQSYA + instrumental → UTILISER DE.

— Introductory prepositions (K, O) are suppressed in titles.

— Russian passive becomes French active if we have Verb + Subject/ARG1.

— In locative phrases, DANS is interpreted as EN or A, according to the class of the noun: 'EN Union Soviétique'.

EFFAC Auxiliary nodes given by TL and not used are deleted (idioms, nonstandard prepositions).

ACTL Particular idiomatic expressions are directly translated:

— infinitive governed by an impersonal + complement:

FORMULU MOZHNO ZAPISATQ → FORMULE PEUT ETRE
NOTEE.

- ESLI+infinitive→SI ON+present.
- 3rd person plural without subject→ON+verb.
- BUDET+infinitive→simple future.
- participle+complement QUI+verb.

QUALD This grammar handles actualization and qualifiers:

- French tenses and modes are computed from their Russian values.
- Pronominal Russian adjectives analysed as epithets become designators.
- Information relative to the determiner is computed on nominal groups according to the type of their designator or of the cardinal, if any.
- The order is changed in nominal groups. In the normal case, adjectives go after the noun, their relative distances being preserved: A1+A2+N →N+A2+A1.

ART Any other designator is used to complete the determination information on nominal groups.

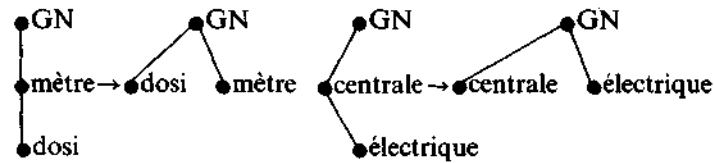
DERV This grammar handles derivations, negation, prefixes and others.

Derivations.

- Verbal adjective or participle without complement: -ANT, -EUR, -E, ...
- Adjective→prepositional group with DE
- Substantive→substantive derived from the adjective: UTILITE, MOTEUR.
- Substantive→substantive derived from the verb (action noun).
- Adverb→adverb derived from the adjective.
- Direct derivation: substantive→substantive, adjective→adjective.

Negation, prefixes and others.

- If there was no negation in Russian, the predicted French one is deleted (NON, PEU, IN).
- Pre- or post-fixes are given in TL under the main node. Here, they are attached at the same level:



- The syntactic function or the subordination conjunction is noted, if this hasn't been done by the analysis.
- The comma preceding a relative clause is deleted.

DTM Determination of nominal groups is computed from the head-node or from its class. For example, family nouns take ABS (no article permitted).

4.2. Examples

ADJ1: 0(\$M, A, \$N, 2)/
 0: K1-E-GN -ØU- K1-E-GP;
 A: \$EPTIT-ET-ADF-E-ADF0-ØU-
 LX-E-PX-ET-K-E-NA;
 2: \$GØV;
 \$M: \$NEPIT -ØU- ADF-E-ADF0/
 ==0(\$M, \$N, 2, A)//.

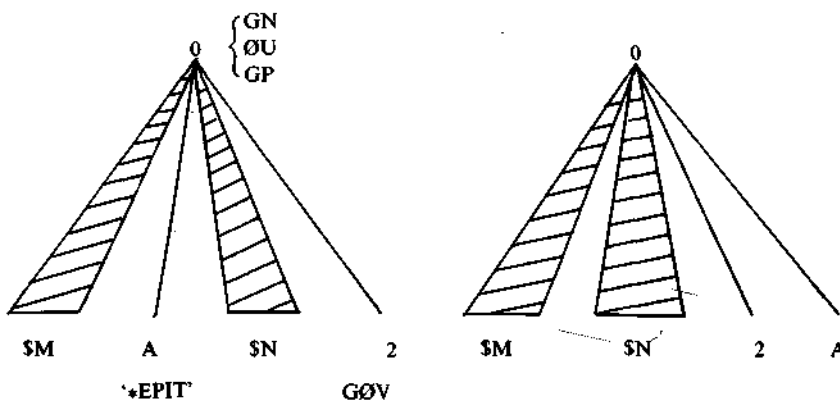


Figure 21.

Rule ADJ1 handles the place of the adjective (horizontal order is significant in the intermediate structure). '\$M' and '\$N' are 'forest-parameters', which have not been explained in section 2. Node 0 dominates forest \$M, node A, forest \$N, node 2 and an unnamed forest, which will be transferred (by default) under new node 0, to the right of new node A. The condition '\$NEPIT -ØU- ADF -E- ADF0' must be verified by all roots of forest \$M. Only universal quantification is provided, as an existential one may readily be expressed by writing one node explicitly.

The effect of two successive applications of rule ADJ1 is shown in Figure 22.

Figure 23 shows the output of this step on the standard example.

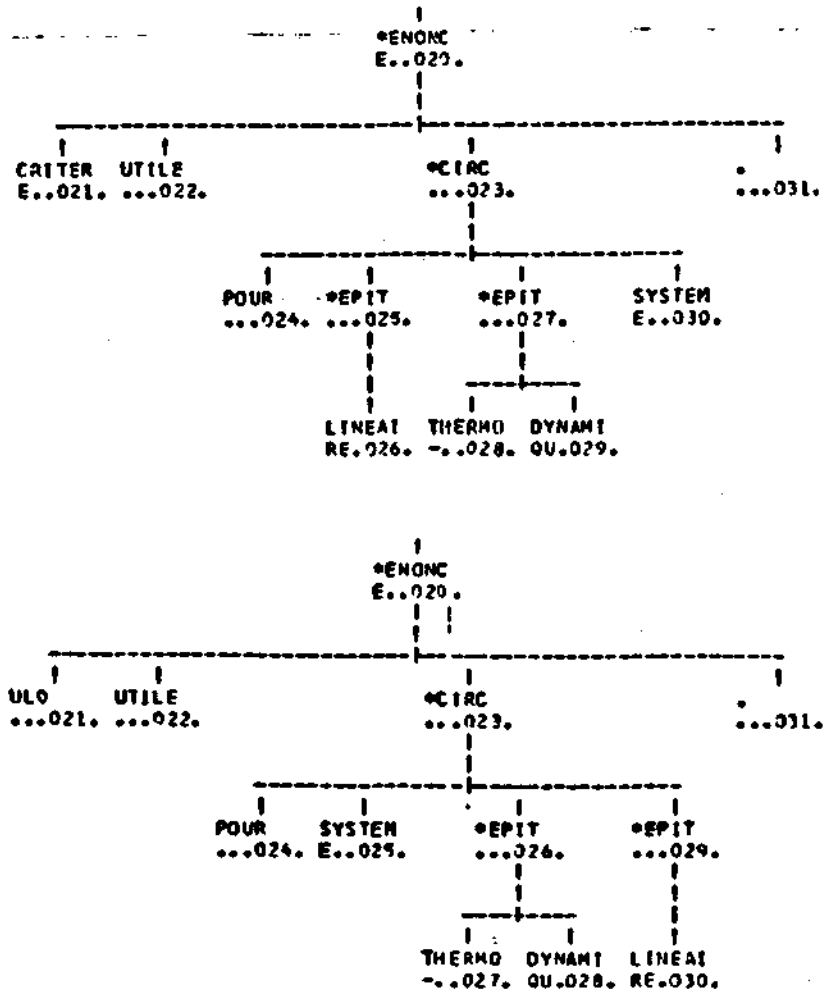


Figure 22.

5. Syntactic generation

ROBRA is also used in this step as algorithmic component (see section 2). The aim of this step is to produce a tree structure where the terminal nodes contain all the information necessary for generating the output text, and express the final surface order of the words. This is a constraint imposed by the nature of the algorithmic component SYGMOR, used in the last step.

5.1. Strategy for the syntactic generation of French

5.1.1. Overall organization. Figure 24 shows the control graph of the TS (transf. system).

5.1.2. Grammars.

RC This grammar copies onto the non-terminal node all variables but ART, FT, ANF, LX from the terminal 'head' (*gouverneur*). It also checks that number and gender have only one value; if not, only one is retained.

AC1 — Coordination of nouns. Information such as determination and preposition (to be repeated) are copied from the main element onto the coordinated terms.

- Place of the subject relative to the verb.
- Generation of a preposition before an infinitive, a gerundive or a prepositional group having values for PRG (see appendix 2).
- Generation of a periphrase when a simple adverbial form does not exist or is not wished (*homogènement* → *de façon homogène*)

ADJ This grammar handles the agreement in gender and number between nouns, adjectives and articles.

RELATIF Chooses the correct paradigm for the pronoun (DONT, QUI, LEQUEL).

AC2 — Retains only one of the homographs noted under ULOCC, in case they have the same lexical unit.

- Handles the (famous) agreement between an object and its attribute.
- Copies values of number and gender computed by grammar ADJ onto the headnote (terminal node) of the adjectival group.
- Values of gender, number and person are copied from the subject onto the sentence node, and then onto the headnode (verb), as well as onto the participle or adjective depending on the infinitive. For example, *nous pouvons paraître étonnés*.
- Generates the hyphen in compound words.
- Handles noun ellipses: the headnode of the nominal group is suppressed, and only its representant is preserved:



ART Generates the article (LE, UN) of a determined nominal group (variable ART has value DEF or IDF).

SOMMET 1 : UL(*TEXTE),
 SOMMET 2 : UL(*ULFRA),
 SOMMET 3 : *FORMULIROVAN: UL(*ENONCE),K1(PH),K(VB),NBR(SIN),A(P),G(M),N(S),P(3),R(FIN),ABS(A1,A2,SJ),T(PAS),
 C(P1ACC),MD(VR3),
 SOMMET 4 : UL(UL),FT(DEB),
 SOMMET 5 : UL(ON),NBR(SIN),TPNISJA),G(M),P(3),
 SOMMET 6 : UL(FORMULER),LX(GOV),K(FVB),DF(DIR),TF(PRE),MF(IND),AF(II),NBR(SIN),RE,(RFL),G(M),N(S),P(3),R(FIN),
 ABS(A1,A2,SJ),C(P1ACC),
 SOMMET 7 : PRINCP: UL(*OBJET),K1(GM),NAT(OBJ),AG(A2),K(NM),ART(DEF),G(M),N(S),P(3),PG(SOB),ANN(INA),
 SOMMET 8 : UL(PRINCIPLE),LX(GOV),NAT(OBJ),AG(A2),K(NM),DF(DIR),NBR(SIN),G(M),P(3),ANN(INA),
 SOMMET 9 : UL(UL),FT(FIN),
 SOMMET 10 : OPREDELYAETSYA: UL(*ENONCE),K1(PH),AG(A2),K(VB),NBR(SIN),A(II),G(M),N(S),P(3),R(FIN),PG(AIZ),
 ABS(A1,A2,SJ),TF(PRE),C(P1ACC),CP2(PPR),MD(VR3),
 SOMMET 11 : UL(UL),FT(DEB),
 SOMMET 12 : UL(*CIRC),K1(GP),AMF(RLT),K(NM),ART(ABS),G(M),N(S),
 SOMMET 13 : UL(A-L-ATDE),CS(ACT),FT(PP),K(PP),AM(INA),MPC(IDE),
 SOMMET 14 : UL(PRINCIPLE),LX(GOV),AMF(RLT),K(NM),DF(DIR),NBR(SIN),G(M),P(3),
 SOMMET 15 : UL(DI),NBR(SIN),TPNISJA),G(M),P(3),
 SOMMET 16 : UL(DEFINIR),LX(GOV),AG(A2),K(FVB),DF(DIR),TF(PRE),MF(IND),NBR(SIN),G(M),N(S),P(3),R(FIN),
 ABS(A1,A2,SJ),C(P1ACC),
 SOMMET 17 : KRITERIJ: UL(*OBJET),K1(GM),CS(ACT),AG(A2),K(NM),ART(DEF),G(M),N(S),P(3),ANN(INA),C(P1DAT),
 SOMMET 18 : UL(CRITERE),LX(GOV),CS(ACT),AG(A2),K(NM),DF(DIR),NBR(SIN),G(M),P(3),ANN(INA),C(P1DAT),
 SOMMET 19 : UL(UL),FT(FIN),
 SOMMET 20 : PRIGODNYI: UL(*ENDNCE),K1(IND),CS(ACT),AG(A2),K(AQ),NBR(SIN),G(M),N(S),PG(SK),ANN(INA),C(P1PPR),
 SOMMET 21 : UL(UL),AMF(RLT),NBR(SIN),TPN(SJA),G(M),P(3),
 SOMMET 22 : UL(UTILE),LX(GOV),PRG(AJQ),CS(ACT),AG(A2),K(FI),DF(DIR),NBR(SIN),G(M),N(S),ANN(INA),
 SOMMET 23 : SISTEMI: UL(*CIRC),K1(GP),NAT(OBJ),K(NM),ART(DEF),G(F),N(S),P(3),PG(5P),PGA(4DLA),ANN(INA),
 SOMMET 24 : DLYA: UL(PUR),NAT(OBJ),FT(PP),K(PP),K(FV),ANN(INA),
 SOMMET 25 : UL(SYSTEME),LX(GOV),NAT(OBJ),K(NM),DF(DIR),NBR(SIN),G(M),P(3),PGA(4D-A),ANN(INA),
 SOMMET 26 : TERMODINAMIKESKOJ: UL(*EPI),K1(GAI),CS(IRT),K(AQ),G(F),N(S),MD(ADJ),
 SOMMET 27 : UL(THERMODYNAMIQUE),LX(GOV),PRG(AJQ),K(FI),DF(DIR),G(F),N(S),MD(ADJ),
 SOMMET 28 : MELINEJMOJ: UL(*EPI),K1(GAI),CS(QAL),K(AQ),G(F),N(S),MG(MET),MD(ADJ),
 SOMMET 29 : UL(LINEAIRE),LX(GOV),PRG(AJQ),K(FI),DF(DIR),G(F),N(S),MD(ADJ),MG(FIN),
 SOMMET 30 : UL(,),FT(FIN),K(CWG),

Figure 23.

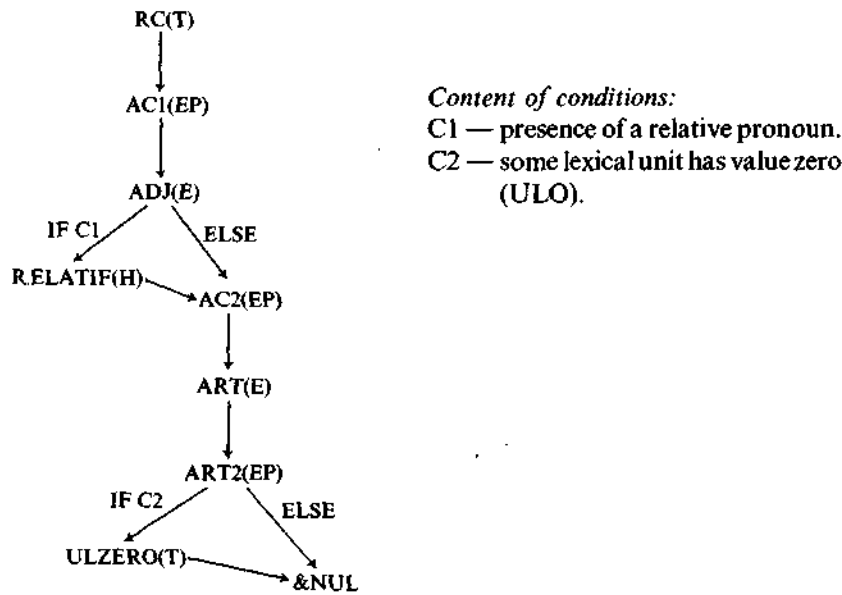


Figure 24.

- ART2** — Controls the place of articles.
 — Generates the reflexive pronoun for verbs in reflexive voice.
 — Generates the correct auxiliary verb for compound tenses or for sentences having an adjective as head.
 — Controls the place of adverbs relative to auxiliary verbs.
 — Sets the subjective mode according to the conjunction or the class of the verb.
 — Generates the negation and puts it in its place (analytical *ne ... pas*, *non-* as prefix, or synthetic, with morphological prefixes IN, IR, etc ...).
 — Generates the degree: PLUS, MOINS.
 — Generates the elements DE, QUE of compound governors: AFIN DE, AFIN QUE.
 — Handles pronouns: QUI/QUE (relative), tonic or atonic form of personal pronouns, synthetic form of pronominal prepositional groups: EN, Y, LUI, LEUR.
 — Generates special punctuation marks to present alternate translations in the output text in cases of unsolved polysemy.

ULZERO Suppresses the nodes where UL=ULO.

5.2. Examples

LEQUEL: 0(1)/
 0: K1-E-GP -ET- ANF-E-RLT;
 1: LX-E-GØV/
 =-0(1)//
 0: 0, ANF:=ANF0, ART:=ABS;
 1: 1,
 -SI- ANM(0)-E-ANI -ALORS- UL(1):='QUI'
 -SINON- UL(1):='LEQUEL' -FSI-

ART1: 0(*, 1)/
 0: (-K1-E-GN-ØU-K1-E-GP) -ET-
 ANF-E-ANFO -ET-
 (-ART-E-DEF-ØU-ART-E-IDF-)
 1: FT-NE-PP -ØU- ART-NE-ABS/
 =-0(2, 1)/2 - */
 0: 0, ART:=ABS; 2: *F, FT:=DT, \$GNB(0),
 -SI- ART(0)-E-IDF -ALØRS- UL(2):='UN'
 -SINØN- UL(2):='LE' -FSI-

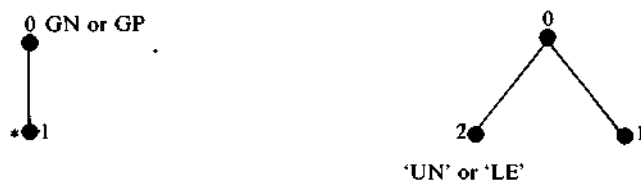


Figure 25.

Rule LEQUEL chooses the appropriate relative pronoun (QUI for an animate referent, LEQUEL otherwise). Rule ART1 generates the appropriate definite or indefinite article. Procedure GNB assigns to it the gender and number of its group (GN or GP). Note the use of the star '*' in the schema to indicate that node 1 is the first son of node 0 (the forest to its left is empty).

The output of this step is shown in Figure 26.

6. Morphological generation

This is the last step of the translation process. Words of the output text are generated. Some facilities must be provided by the algorithmic component to handle elisions and contractions.

SUMMET 1 : UL(*TEXTEI).
 SUMMET 2 : UL(*ULPRA).
 SUMMET 3 : SFORMULIRJAVAN: UL(*ENONCEI,KF(VB),KI(PIH),DF(DIR),TF(PREI),MF(END),AF(LI),NBR(SIN),TRC(RCH),P(3),G(M),
 PF(N)).
 SUMMET 4 : UL(OR),NBR(SIN),TPN(SJA),P(3),G(M).
 SUMMET 5 : UL(AVJ),FT(AK),KF(VB),TF(PREI),MF(TND),NBR(SIN),P(13).
 SUMMET 6 : UL(PDRMUR),KF(VB),LX(GOV),DF(DIR),MF(PRC),NBR(SIN),TRC(RCB),G(M),RF(IN).
 SUMMET 7 : PRINCIP: UL(*OJJEI),AG(A2),KF(INM),KI(GNI),DF(DIR),ART(ABS),NBR(SIN),TRC(RCH),P(3),G(M),ANM(INA).
 SUMMET 8 : UL(LE),FT(OT),NBR(SIN),G(M).
 SUMMET 9 : UL(PJACTIPEI),AG(A2),KF(INM),LX(GOV),DF(DIR),NBR(SIN),P(3),G(M),ANM(INA).
 SUMMET 10 : OPREDELYAETSYA: UL(*ENONCEI,AG(AG),KF(VB),KI(PIH),DF(DIR),TF(PREI),MF(IND),NBR(SIN),TRC(RCH),P(3),G(M),
 PF(N)).
 SUMMET 11 : UL(*CIRC),KF(INM),KI(GPI),DF(DIR),ART(ABS),NBR(SIN),TRC(RCH),P(3),G(M).
 SUMMET 12 : UL(L-A-TOE),FT(PP),ANM(INA).
 SUMMET 13 : UL(DE),FT(PP).
 SUMMET 14 : UL(LF),LX(L),KF(INM),LX(SOV),ANF(ALT),DF(DIR),NBR(SIN),P(3),G(M).
 SUMMET 15 : UL(ON),NBR(SIN),TPN(SJA),P(3),G(M).
 SUMMET 16 : UL(DEFINIA),AG(A6),KF(VB),LX(GOV),DF(DIR),TF(PREI),MF(IND),NBR(SIN),TRC(RCB),P(3),G(M),RF(IN).
 SUMMET 17 : KRITERIJ: UL(*OBJET),AG(A2),KF(INM),KI(GNI),DF(DIR),ART(ABS),NBR(SIN),TRC(RCH),P(3),G(M),ANM(INA).
 SUMMET 18 : UL(LE),FT(OT),NBR(SIN),G(M).
 SUMMET 19 : UL(CITERI),AG(A2),KF(INM),LX(GOV),DF(DIR),NBR(SIN),P(3),G(M),ANM(INA).
 SUMMET 20 : PHIGODVITJ: UL(*ENONCE),AG(A6),KF(AQ),KI(MD),DF(DIR),NBR(SIN),PRG(AJQ),TRC(RCB),P(3),G(M),ANM(INA).
 SUMMET 21 : UL(LITTLE),AG(A6),KF(AQ),LX(GOV),DF(DIR),NBR(SIN),PRG(AJQ),P(3),G(M),ANM(INA).
 SUMMET 22 : SISTEMYI: UL(*CIRC),KF(INM),KI(GPI),DF(DIR),ART(ABS),NBR(SIN),TRC(RCH),P(3),G(M),ANM(INA).
 SUMMET 23 : ULVA: UL(PJAK),FT(PP),KF(LIV),ANM(INA).
 SUMMET 24 : UL(LE),FT(OT),NBR(SIN),G(M).
 SUMMET 25 : UL(SYSTEME),KF(INM),LX(GOV),DF(DIR),NBR(SIN),P(3),G(M),ANM(INA).
 SUMMET 26 : Tsa(MO)INAMIKESKOJ: UL(PEPITI),KF(A3),KI(GA),DF(DIR),NBR(SIN),PRG(AJQ),TRC(RCB),P(3),G(M).
 SUMMET 27 : UL(THA(MO)DYNAMIQUEI),KF(AQ),LX(GOV),DF(DIR),NBR(SIN),PRG(AJQ),TRC(RCB),P(3),G(M).
 SUMMET 28 : NEL(INF)MOJ: UL(PEPITI),KF(AQ),KI(GA),DF(DIR),NBR(SIN),PRG(AJQ),TRC(RCB),P(3),G(M),NGF(NON).
 SUMMET 29 : UL(NON),LX(PX).
 SUMMET 30 : UL(LINEAIKE),KF(AQ),LX(GOV),DF(DIR),NBR(SIN),PRG(AJQ),TRC(RCB),P(3),G(M).
 SUMMET 31 : UL(LE),FT(PP).

Figure 26.

6.1. The algorithmic component: SYGMOR

This model is a composition of two transducers: the first, 'tree-to-string', produces the frontier of the object tree; the second transforms this string (of masks of variables) into a string of characters, under the control of the linguistic data. These data are made up of:

- declaration of variables, formats as well as condition and assignment procedures.
- dictionaries (with direct addressing by the values of certain declared variables). The first dictionary must be referenced by the UL.
- a grammar.

Each item in a dictionary gives a list of <condition/assignment/string> triplets, the last one having an empty (NIL) condition.

```

A-L-AIDE== /VID/'A L'AIDE
AVOIR ==TP1A /VID/'AI,
      ==PSSPT/V3H/'EU,
      ==TP3A /V3A/'A.
LEQUEL ==N1B /VID/LAQUELLE,
       ==N1D /VID/'LESQUELLES,
       ==PLU /VID/'LESQUELS,
       == /VID/'LEQUEL.

```

Figure 27.

TP1A, PSSPT, TP3A are names of condition procedures. VID, V3H, V3A are names of formats. The apostrophes ('AI) are used in the grammar to make contractions.

At any point during processing, certain quantities are accessible and are referenced in the metalanguage of the grammar by special symbols:

- '1', '2', ..., 'n' are the numbers of the dictionaries ($n \leq 8$).
- 'C' and 'P' are the 'current' and 'preceding' masks.
- 'T' and 'S' are two string variables, T being the one currently processed and S the one previously output. They are also names of two associated masks of variables.
- 'G', 'D' and 'M' are three origins, left, right and middle, in T. They define the points where a string (given by a dictionary item) may be inserted. Note that it would have been equivalent to define two string variables for the 'left part' and 'right part' of T.

A grammar rule is made of:

- a condition of application for the accessible masks and strings.
- a part for reading in one or more dictionaries and manipulating the accessible strings.

- an assignment part, where values are given to the C, P, T masks, using the dictionaries, and to the P and S masks, using C and T.
- a 'string transformation part' to realize replacements of substrings in T.
- finally, a part indicating a sequence of transitions to follow after the application of the rule. The rules between parentheses are optional.

Example:

```

ELS3: DD(C)-NE-DD4-ET-DD(C)-NE-DD0-ET-DD(P)-E-DD4
-OU-DD(C)-NE-DD4-ET-DN(C)-NE-DN0-ET-DD(P)-E-DD4
==/T:=S/
TCHaine(M, -4, "LE'", "L'");
TCHaine(M, -4, "LA'", "L'");
TCHaine(M, -3, "DU'", "DE L'");
TCHaine(M, -3, "AU'", "A L'");
TCHaine(M, -4, "MA'", "MON");
TCHaine(M, -4, "TA'", "TON");
TCHaine(M, -4, "SA'", "SON");
TCHaine(M, -5, "CET'", "CET");
TCHaine(M, -5, "CET'", "CE");
TCHaine(M, -5, "EAU'", "EL");
TCHaine(M, -4, "OU'", "OL");
TCHaine(M, -7, "VIEUX'", "VIEIL");
TCHaine(M, -8, "TOUTES'", "TOUT");
TCHaine(M, -7, "TOUTE'", "TOUT");
TCHaine(M, -4, "SI'IL", "S'IL");
TCHaine(M, -4, "DE'LES", "DES");
TCHaine(M, -4, "DE'LE", "DU");
TCHaine(M, -3, "A'LES", "AUX");
TCHaine(M, -3, "A'LE", "AU");
TCHaine(M, -2, "'L", "L");
TCHaine(M, -3, "E'", "");
TCHaine(M, -2, "", "");
TCHaine(M, -2, "", "");
TCHaine(M, -1, "", "")/.
    
```

Figure 28.

Rule ELS3 handles almost all cases of contraction in French. The output string S is assigned to the current string T. The substitution operations (TCHaine) are applied in sequence. For instance, JE' 'AI will be transformed into J'AI by 'TCHaine(M,-3,"E' ','"'). -3 gives the origin relative to M.

It should be noted that, unlike ATEF, SYGMOR realizes a finite-state deterministic automaton, thus reflecting the lesser complexity of the synthesis process. To process a mask, SYGMOR looks for the first applicable rule (at least one must have an empty condition), applies it and

follows the transitions indicated, unless it finds an inapplicable obligatory rule. In this case, the system executes the special rule MOTINC, or a default action if this rule has not been declared. It is thus possible to generate an arbitrary error indicator at that point. For instance, non-translated source lexical units will be printed between special markers.

6.2. Linguistic strategy for the synthesis of French morphemics

The first applicable rule assigns to T the 'initial chain' (IC) given by the dictionary. Other rules are used to generate the correct surface form, according to the information contained in mask C, and to take care of elisions and contractions.

1. Preterite or subjunctive imperfect

The initial chain is modified for verbs like FINIR (IC=FINISS), where the SS is suppressed. The endings are found in dictionary 2; they are of 4 types: PARL-A, FINI-T, VEND-IT, COUR-UT. In the first type, GA becomes GEA (MANG-E-A).

2. Past participle or homograph action noun

The IC is modified for verbs like FINIR. There are 6 paradigms: PARL-E, SORT-I, CRAIN-T, FINI-#, COUR-U, MI-S. If the IC is special for the participle, it will be handled as the corresponding adjective: NE, ABSOUS.

3. Gerundive — Present participle

This is directly constructed in the grammar by adding ANT to the IC.

4. Indicative present singular

There are 3 paradigms: PARL-E, ROMP-T, VEND-#. The following transformations are performed:

- a. BROY+E→BROIE; JET+E→JETTE; APPEL+E→APPELLE,
- b. FINISS+T FINIT (as for PARAISS, CROISS).

5. Third person plural of present subjunctive or indicative, or singular for present subjunctive

There is only one paradigm: E, ES, ENT. Transformations a. of 4. above are performed.

6. First or second person plural of present indicative, subjunctive and imperative. All imperfect forms.

There is only one paradigm: ONS, EZ, AIS, AIT, AIENT, IONS, IEZ. 'GO' and 'GA' become 'GEO' and 'GEA' (MANGEONS).

7. Infinitive

There are 6 paradigms: ER, RE, R, IR, DRE, OIR. The IC of verbs like FINIR is transformed (FINISS+R→FINIR).

8. *Simple future or conditional present*

There are four paradigms: ERA, IRA, RA, DRA. The transformations are the same as for present indicative singular (4.).

9. *Action noun*

a. If irregular, it is given by the IC, without suffixation. Only double final consonants are simplified: FUSIONN→FUSION; DONN→DON; ESSAY→ESSAI; PROGRESS→PROGRES; GUETTE→GUET; TRAVAILL→TRAVAIL.

b. In the regular case, the following transformations are performed:

— doubling of T, L before E: CACHET→CACHETTE, ETINCEL→ETINCELLE.

— Y→I: PAY→PAIE.

— for verbs having value AT for variable IFX, 'AT' is inserted between the root and the suffix ION, URE: OPER-AT-ION. Transformations are performed: QU→C, GU→G (*navigation, troncature*).

— if the root ends with a G, an E is added before A, U: VENG-E-ANCE, GAG-E-URE.

10. *Agent noun or adjective*

The following transformations are performed:

— AT is inserted before EUR: OPER-AT-EUR, but FABRIC-ANT.

— GA becomes GEA: ENCOURAG-E-ANT.

11. *Potential adjective and its abstract noun*

a. If the derivation is irregular, see adjective: SOLUBL.

b. If regular, it may be ABLE or IBLE. The following transformations are performed:

QU-A→CA; GU-A→GA; G-A→GEA:
EXPLIC-ABLE; NAVIG-ABLE, NEGLIG-E-ABLE.

12. *Abstract noun derived from the adjective*

If irregular, it is given directly by the dictionary. See noun (14). If regular, the transformations vary according to the suffix:

— for ITE: IER→AR (*particularité*); AIR→AR (*parité*); AIN→AN (*vanité*); ELL→AL (*formalité*); EUS→OS (*porosité*); BL→BIL (*probabilité*).

— for ISME: AIR→AR (*sectarisme*); BL→BIL (*misérabilisme*); ISTE→# (*probabilisme*), IQUE→# (*civisme*).

— for NCE: the final T disappears (*urgence*).

13. *Adverb derived from the adjective*

If irregular, this is given directly as an IC in the dictionary. If not,

EMENT is added to the root, with the transformations:

— NT→MM PATIEMENT.

— E is suppressed after a vowel: JOLIMENT, CONTINUMENT.

14. *Non-derived noun or adjective*

If the form is invariant, see above. If not, the process is as follows:

a. *Adjectival paradigms.* The root is the feminine form. To generate the masculine, the following transformations are performed:

— EUS, RIC, ERESS→EUR: FARCEUR, ACTEUR, ENCHANTEUR.

— ESS→E: POETE

— SS→S: GRAS

— SS, S→X: ROUX, HEUREUX

— simplification of TT, NN, LL: COQUET, BON, MANUEL.

— V→F: VEUF

— QU→C: TURC.

For the masculine plural, AL, AIL→AUX: NORMAUX, BAUX. If the root is the same for masculine and feminine, an E is added: UTILE.

b. *Plural of nouns is S and of adjectives.*

— S is added if there is no S or X already present.

15. *Prefixation, elision, contraction*

a. *The synthetic negation* by IN is created if variable NGF is not zero. The transformations are: IN+L→IL (*illisible*); IN+M, B, P→IM (*improbable*); IN+R→IR (*irréel*).

b. *Elision*

— Prepositions and conjunctions in -E, as well as LE, LA, lose the vowel before “'”, a special sign given in the dictionaries to indicate that the root begins with a vowel or a silent ‘h’.

— Also: MA, TA, SA→MON, TON, SON

BEAU, NOUVEAU, FOU, MOU, VIEUX→BEL,
NOUVEL, FOL, MOL, VIEIL.

— Before a consonant, CET becomes CE.

— TOUTES, TOUTE as variable adverbs before a vowel become TOUT.

— If the “'” remains in other positions, it is erased.

c. *Contraction*

— DE+LE→DU; DE+LES→DES; A+LE→AU; A+LES→AUX (same with LEQUEL); SI+IL→S’IL.

— Punctuation marks are concatenated with the preceding form, except for the opening parenthesis, which is concatenated with the following form. Hyphens (–) and dashes (—) are concatenated with both (normally, SYGMOR inserts a space between two successive forms).

Received 1 October 1980
Revised version received
13 July 1981

Groupe d'Études pour la
Traduction Automatique
Bâtiment C.E.T.A.
Domaine Universitaire
38041 Grenoble, CEDEX 53
France

Appendix 1

Examples of translations

The following is text 'AVIA' in corpus 'RUSSE'. In ARIANE-78, corpora are collections of texts sharing the same list of hierarchical separators (for chapter, section, paragraph, and perhaps sentence). Texts are automatically segmented into 'translation units' of parametrizable size. Each translation unit is a sequence of consecutive logical units of the highest possible level, with respect to the bounds given on its size. For example, a translation unit may have between 40 and 80 words, or between 120 and 200 words ..., these parameters being set depending on the available virtual memory.

In this example, we see the date and time of output (July 18th), the date and time of translation (June 26th), as well as the dates when the linguistic data used in this translation were compiled. 'A' stands for analysis (AM-AS), 'T' for transfer (TL-TS), and 'G' for generation (GS-GM).

The text itself is taken from a (real) article on aeronautics and is more of a literary kind than purely technical. Note the denotation of alternate translations by '?'. For example, MIR is a famous ambiguity between 'world' and 'peace' (MIR MIRU). Some mistakes are due to bugs in one of the grammars (verb + direct object + subject has been forgotten, hence the construction: *ont obtenu ... les lignes aériennes*). More serious problems arise with articles, and determination in general. Russian has no articles, and the use of them in French is far from clear.

RUSSE AVIA

18 JUILLET 1979 10H 56MN 51S

LANGUES DE TRAITEMENT: RUS-FRA

TEXTE D'ENTREE:

RABOTNIKI GRAZHDANSKOJ AVIACII
SAMOOTVERZHENNYIM TRUDOM VNOSYAT OGNOMNYI
VKLAD V USPESHNOE VYIPOLNENIE PLANOV NARODNOGO

XOZYAJSTVA, V RESHENIE DALQNEJSHIX ZADAKH
KOMMUNISTIKESKOGO STROITELQSTVA. PLAN PЕРVOGO
GODA PYATILETKI V OBLASTI VOZDUSHNOGO
TRANSPORTA PEREVYIPOLNEN. ZNAKHITELQNO VYIROSЛИ
OBWEMYI RABOT, VYIPOLNYAEMYIE AVIACIEJ V
NARODNOM XOZYAJSTVE. POLUKHILI DALQNEJSHEE
RAZVITIE MEZHDUNARODNYIE VOZDUSHNYIE SVYAZI. V
NASTOYATHEE VREMYA, ZAKLYUKHENYI PYATQ NOVYIX
SOGLASHENIJ OB AVIACIONNOM SOOBTHENII S
ZARUBEZHNYIMI STRANAMI. OTKRYITYI SEMQ NOVYIX
MEZHDUNARODNYIX AVIALINIJ. SOVETSKIE SAMOLETYI
SEGODNYA SOVERSHAYUT REGULYARNYIE POLETYI
PRAKTIKHESKI VO VSE STRANYI MIRA. MINISTERSTVOM
GRAZHDANSKOJ AVIACII RAZRABOTAN KOMPLEKSNYIJ
PLAN MEROPRIYATIJ PO POVYISHENIYU YEFFEKTIVNOSTI
PROIZVODSTVA I KAKHESTVA RABOTYI. V NEM
IZLOZHENYI VAZHNYIE NAPRAVLENIYA NASHEJ
DEYATELQNSTI, PREDUSMOTRENYI PUTI
MAKSIMALQNOGO ISPOLQZOVANIYA VNUTRENNIX
REZERVOV I OBESPEKHENIYA BEZOPASNOSTI POLETOV.
OSOBOE VNIMANIE UDELENO POVYISHENIYU TEMPOV
ROSTA PROIZVODITELQNSTI TRUDA I
PROFESSIONALQNOGO MASTERSTVA INZHENEROV I
TEKNIKOV.

TEXTE DE SORTIE:

----- (TRADUCTION DU 26 JUIN 1979 18H 19MN 12S) -----
VERSIONS: (I: 22/06/79; T: 25/06/79; G: 25/06/79)

LES TRAVAILLEURS DE L'AVIATION CIVILE PAR LE TRAVAIL
DEVOUE APPORTENT LA CONTRIBUTION ENORME DANS LA
REALISATION REUSSIE DES PLANS DE L'ECONOMIE
POPULAIRE, DANS LA RESOLUTION DES PROBLEMES
ULTERIEURS DE LA CONSTRUCTION COMMUNISTE. LE
PLAN DE LA PREMIERE ANNEE DU QUINQUENNAT DANS LE
DOMAINE DU TRANSPORT AERIEN EST DEPASSE. ONT
CONSIDERABLEMENT ACCRU LES VOLUMES DES TRAVAUX
REALISES PAR L'AVIATION DANS L'ECONOMIE POPULAIRE.
ONT OBTENU LE DEVELOPPEMENT ULTERIEUR LES
LIAISONS AERIENNES INTERNATIONALES. A PRESENT, ON A
CONCLU CINQ ACCORDS NOUVEAUX DE LA
COMMUNICATION D'AVIATION AVEC LES PAYS

ETRANGERS. ON A OUVERT? DECOUVERTES? SEPT LIGNES AERIENNES INTERNATIONALES NOUVELLES. LES AVIONS SOVIETIQUES EFFECTUENT AUJOURD'HUI LES VOLS REGULIERS PRATIQUEMENT DANS TOUS LES PAYS DU MONDE? PAIX?. LE MINISTERE DE L'AVIATION CIVILE A ELABORE LE PLAN COMPLEXE DES MESURES SUR L'AUGMENTATION DE L'EFFICACITE DE LA PRODUCTION ET DE LA QUALITE DU TRAVAIL. ON Y A EXPOSE LES DIRECTIONS IMPORTANTES DE NOTRE ACTIVITE, ON A PREVU LES VOIES DE L'UTILISATION MAXIMALE DES RESERVES INTERIEURES ET DE L'ASSURANCE DE LA SURETE DES VOLS. L'ATTENTION SPECIALE EST ACCORDEE A L'AUGMENTATION DES RYTHMES DE LA CROISSANCE DE LA PRODUCTIVITE DU TRAVAIL ET DE LA MAITRISE PROFESSIONNELLE DES INGENIEURS ET DES TECHNICIENS.

The next text is syntactically simpler than the preceding one, although not really technical either.

Note the realization of the negation by prefixing with NON (non-conversation, non-axial) *Dans le mot d'entrée* (in the word of introduction) could be better translated by *dans l'introduction*. To obtain it, one should put it as an idiom in the dictionary (TL).

RUSSE RAPPORT

18 JUILLET 1979 10H 58MN 51S

LANGUES DE TRAITEMENT: RUS-FRA

TEXTE D'ENTREE:

SIMPOZIUM POSVYATHEN YADERNOJ SPEKTROSKOPII I STRUKTURE ATOMNOGO YADRA. VO VSTUPITELQNM SLOVE PODKHERKIVAETSYA VAZHNAJA ROLQ, KOTORUYU SIMPOZIUM SYIGRAL V RAZVITII YADERNOJ FIZIKI SLABYIX YENERGIJ V SOVETSKOM SOYUZE. V OXODE SIMPOZIUMA OBSUZHDEN RYAD VAZHNYIX ISSLEDOVANIJ, OSUTHESTVLENNYIX SOVETSKIMI UKHENYIMI. V KHASTNOSTI, IZUKHENO NESOXRANENIE KHETNOSTI V YADERNYIX PROCESSAX, SOZDANIE MODELI NEAKSIALQNOGO YADRA, SPONTANNOE DELENIE IZOTOPOV SVERXTYAZHELYIX YELEMENTOV I OBNARUZHENIE YEFFEKTA TENEJ PRI RASSEYANII

KHASTIC. SOBRANYI UBEDITELQNYIE STATISTIKHESKIE
DANNYIE, OTRAZHAYUTHIE ROST KHISLA
PREDLOZHENNYIX DOKLADOV. OTMEKHAETSYA
PRISUTSTVIE SREDI UKHASTNIKOV SPECIALISTOV IZ
ZARUBEZHNYIX STRAN.

TEXTE DE SORTIE:

----- (TRADUCTION DU 26 JUIN 1979 17H 59MN 23S) -----
VERSIONS: (A: 22/06/79; T: 25/06/79; G: 25/06/79)

LE SYMPOSIUM EST CONSACRE A LA SPECTROSCOPIE
NUCLEAIRE ET A LA STRUCTURE DU NOYAU ATOMIQUE.
DANS LE MOT D'ENTREE ON SOULIGNE LE ROLE
IMPORTANT QUE LE SYMPOSIUM A JOUE DANS LE
DEVELOPPEMENT DE LA PHYSIQUE NUCLEAIRE DES
ENERGIES FAIBLES EN UNION SOVIETIQUE. PENDANT LE
SYMPOSIUM ON A EXAMINE LA SERIE DES ETUDES
IMPORTANTES REALISEES PAR LES SAVANTS SOVIETIQUES.
EN PARTICULIER, ON A ETUDIE LA NON-CONSERVATION DE
LA PARITE DANS LES PROCESSUS? PROCEDES? NUCLEAIRES,
CREATION DU MODELE DU NOYAU NON-AXIAL, DIVISION
SPONTANEE DES ISOTOPES DES ELEMENTS SUPERLOURDS
ET DECOUVERTE DE L'EFFET DES OMBRES POUR LA
DISPERSION DES PARTICULES. ON A REUNI LES DONNEES
STATISTIQUES CONVAINCANTES QUI REFLETENT LA
CROISSANCE DU NOMBRE DES RAPPORTS PROPOSES. ON
REMARQUE LA PRESENCE PARMIS LES PARTICIPANTS DES
SPECIALISTES DES PAYS ETRANGERS.

In the following text 'RUSSE CIOL', the proper noun CIOLKOVSKIJ was not in the dictionary. The morphological generation brackets it by special signs '<' and '>'. 'DE' keeps the quote which indicates a possible elision, due to a default in the grammar. Another bug (*résolution organiser*) comes from an error in the dictionary, where the syntactic government should be the same for the substantive as for the verb.

RUSSE CIOL

18 JUILLET 1979 10H 57MN 56S

LANGUES DE TRAITEMENT: RUS-FRA

TEXTE D'ENTREE:

ORGANIZACIONNIY KOMITET KHTENIJ PRINYAL
RESHENIE ORGANIZOVATQ YETOT SIMPOZIUM V RAMKAX
EZHEGODNYIX NAUKHNYIX KHTENIJ. SIMPOZIUM
ZASLUSHAL OKOLO DVADCATI DOKLADOV, KOTORYIE
DAYUT PREDSTAVLENIE O KONKRETNYIX PROBLEMAX I
METODAX PROGNOZIROVANIYA RAZVITIYA
KOSMONAVTIKI. UKHASTNIKI SIMPOZIUMA
PODKHERKNULI OGROMNUYU ROLQ NAUKHNOGO
NASLEDIYA CIOLKOVSKOGO. OTMEKHAETSYA GLUBINA
RAZRABOTKI CIOLKOVSKIM RAZLIKHNYIX PROGNOZOV.
DOKLADYI NAPRAVLENYI NA TEORETIKHEKUYU I
PRAKTIKHEKUYU RAZRABOTKU PROGNOSTIKHEKIX
IDEJ CIOLKOVSKOGO. ONI SPOSOBSTVOVALI PROPAGANDE
ZADAKH NAUKHNOGO PROGNOZIROVANIYA.
VVIYAVLENYI NOVYIE PROBLEMYI, VSTAYUTHIE PERED
PROGNOZIROVANIEM. UKREPLYAETSYA PRIORITET
SOVETSKOGO SOYUZA V OBLASTI PROGNOZOV PO
RAZLIKHNYIM NAPRAVLENIYAM.

TEXTE DE SORTIE:

----- (TRADUCTION DU 7 MAI 1979 19H 01MN 22S) -----
VERSIONS: (A: 8/03/79; T: 5/05/79; G: 5/05/79)

LE COMITE D'ORGANISATION DES LECTURES A PRIS () LA
RESOLUTION ORGANISER CE SYMPOSIUM DANS LE CADRE
DES LECTURES SCIENTIFIQUES ANNUELLES. LE SYMPOSIUM
A ENTENDU PRES DE VINGT RAPPORTS QUI DONNENT LA
REPRESENTATION DES PROBLEMES CONCRETS ET LES
METHODES DU PRONOSTIC DU DEVELOPPEMENT DE
L'ASTRONAUTIQUE. LES PARTICIPANTS DU SYMPOSIUM
ONT SOULIGNE LE ROLE ENORME DE L'HERITAGE
SCIENTIFIQUE DE' <CIOLKOVSKIJ>. SE REMARQUE LA
PROFONDEUR DE L'ELABORATION PAR <CIOLKOVSKIJ> DES
PRONOSTICS DIFFERENTS. LES RAPPORTS SONT DIRIGES
SUR L'ELABORATION THEORIQUE ET PRATIQUE DES IDEES
DE PRONOSTIC DE' <CIOLKOVSKIJ>. IL CONTRIBUAIT A LA
PROPAGANDE DES PROBLEMES DU PRONOSTIC
SCIENTIFIQUE. ON A REVELE LES PROBLEMES NOUVEAUX
QUI SE LEVENT DEVANT LE PRONOSTIC. ON RENFORCE LA
PRIORITE DE L'UNION SOVIETIQUE DANS LE DOMAINE DES
PRONOSTICS EN DIRECTIONS DIFFERENTES.

This translation is older than the preceding ones, and alternate translations appear between parentheses. '(EXPR)' translates "E2, which notes a formula of some kind, taking half a line of text (2/4). "E3 would take 3/4 of a line, and so on. This has been improved since.

De sorte non-former should be de sorte à ne pas former. This is due to an error in handling negation.

DRME1 L4110

18 JUILLET 1979 11H 9MN 35S

LANGUES DE TRAITEMENT: RUS-FRA

TEXTE D'ENTREE:

SL.110. POSTROENIE VIRIALQNOGO RAZLOZHENIYA S POMOTHQYU KANONIKHESKOGO PREOBRAZOVANIYA. "BIBLIO. GAMILQTONIAN SISTEMYI VZAIMODEJSTVUYUTHIX KHASTIC PREOBRAZUETSYA S POMOTHQYU KANONIKHESKOGO PREOBRAZOVANIYA, VYIDELYAYUTHEGO IZ NEGO "E2-KHASTIKHNYIE SOSTOYANIYA. POSLE KANONIKHESKOGO PREOBRAZOVANIYA VYIDELENNYIE SOSTOYANIYA ZAPISYIVAYUTSYA V VIDE SYMMYI SOOTVETSTVUYUTHIX IDEALQNYIX GAZOV, A POTENCIAL VZAIMODEJSTVIYA ISXODNYIX KHASTIC PERENORMIRUETSYA TAKIM OBRAZOM, KHTOBYI NE OBRAZOVYIVATQ UKHTENNYIX SOSTOYANIJ. STROITSYA TERMODINAMIKA PREOBRAZOVANNOJ SISTEMYI V PERVOM PORYADKE TEORII VOZMUTHENIYA V SOKHETANII S VARIACIONNYIM PRINCIPOM. POKAZANO, KHTO PREDLAGAEMAYA PERESTROJKA GAMILQTONIANA OKAZYIVAETSYA TERMODINAMIKHESKI VYIGODNOJ. POSLE KANONIKHESKOGO PREOBRAZOVANIYA TERMODINAMIKHESKIJ POTENCIAL ZAPISAN V VIDE RAZLOZHENIYA TIPA VIRIALQNOGO. AVTOREFERAT.

TEXTE DE SORTIE:

----- (TRADUCTION DU 23 FEVRIER 1979 17H 47MN 10S) -----
VERSIONS: (A: 22/02/79; T: 23/02/79; G: 23/02/79)

<SL.110.> CONSTRUCTION DU DEVELOPPEMENT (DECOMPOSITION) VIRIEL A L'AIDE DE LA TRANSFORMATION CANONIQUE. (BIBLIO). L'HAMILTONIEN

DU SYSTEME DES PARTICULES D'INTERACTION EST TRANSFORME A L'AIDE DE LA TRANSFORMATION CANONIQUE QUI EN ISOLE LES ETATS A (EXPR) PARTICULES. APRES LA TRANSFORMATION CANONIQUE LES ETATS ISOLEES SONT NOTES SOUS LA FORME DE LA SOMME DES GAZ IDEAUX CORRESPONDANTS, MAIS LE POTENTIEL DE L'INTERACTION DES PARTICULES INITIALES EST RENORMALISE DE SORTE NON-FORMER DES ETATS CONSIDERES. ON CONSTRUIT LA THERMODYNAMIQUE DU SYSTEME TRANSFORME DANS LE PREMIER ORDRE DE LA THEORIE DE LA PERTURBATION EN LIAISON AVEC LE PRINCIPE DE VARIATION. EST INDIQUE QUE LA RECONSTRUCTION PROPOSEE DE L'HAMILTONIEN S'AVERE THERMODYNAMIQUEMENT AVANTAGEUSE. APRES LA TRANSFORMATION CANONIQUE LE POTENTIEL THERMODYNAMIQUE EST NOTE SOUS LA FORME DU DEVELOPPEMENT (DECOMPOSITION) DU TYPE DE LA VIRIELLE. RESUME.

In text 'ESRUS TEST', we see that 'expressions' have been translated by leaving enough space to fill after translation with the corresponding formulae. A minor error occurs with *bien elle s'accorde*, which should be *elle s'accorde bien*.

Note that all these outputs are taken directly from the machine, without any human revision. This level of quality is certainly high enough not to give too heavy a burden to revisors. Nevertheless, the main problem lies in preparing large linguistic data (grammars, dictionaries) and keeping the quality homogeneous over a large corpus. An operational MT system, in our view, can give 'good enough' (not 'perfect') outputs ONLY when tuned and upgraded on a LARGE CORPUS WITH RESTRICTED DOMAIN AND HOMOGENEOUS STYLE.

ESRUS TEST

18 JUILLET 1979 11H 1 MN 29S

LANGUES DE TRAITEMENT: RUS-FRA

TEXTE D'ENTREE:

"E1-PEREXOD S UROVNYA "E2 IDENTIFICIROVALSYA
PUTEM IZMERENIJ "E1-SOVPADENIJ V KASKADE "E1.
RADIACIONNAYA SHIRINA DLYA PEREXODA "E2

OKAZALASQ RAVNOJ "E2. POLUKHENNOE ZHAKHENIE NIZHE, KHEM DAYUT RASKHETYI PO MODELI OBOLOKHEK S "E1-SVYAZQYU, NO XOROSHO SOGLASUETSYA S PREDSKAZANIYAMI ROTACIONNOJ I "E1-MODELEJ DLYA "E1-POLOSJI.

TEXTE DE SORTIE:

----- (TRADUCTION DU 25 JUIN 1979 18H 11MN 39S) -----
 VERSIONS: (A: 22/06/79; T: 25/06/79; G: 25/06/79)

LE PASSAGE ----- DU NIVEAU ----- ETAIT IDENTIFIE AU MOYEN DES MESURES DES COINCIDENCES ----- DANS LA CASCADE ----- LA LARGEUR DE RADIATION POUR LE PASSAGE ----- S'EST AVEREE EGALE A ----- LA VALEUR OBTENUE EST PLUS BASSE QUE DONNENT LES CALCULS EN MODELE DES ENVELOPPES AVEC LA LIAISON -----, MAIS BIEN ELLE S'ACCORDE AVEC LES PREDICTIONS DES MODELES DE ROTATION ET ----- POUR LA ZONE -----

Appendix 2

Variables RUS-FRA

Level	Type	AM	AS	TL	TS	GS	GM	Names and values	Comments
C	E	*	*	*	*	*	*	K=(NM, VB, AQ, AV, PP, CJ, CB, IV, VG, NA, RB)	MORPHOLOGICAL CLASS: noun, verb, adjective+participle, adverb, preposition, coordinating and subordinating conjunctions, particle, punctuation, unalphabetical form, bibl. reference sign.
A	N	*	*					C=(1, 2, 3, 4, 5, 6, 7)	CASE: nominative, inanimate and animate accusative, genitive, dative, prepositional, instrumental

Key: level: P — paradigmatic (if t — tactical), C — categorial, S — syntagmatic, V — valency (government), A — actualization, F — functional, R — argumentary, L — lexico-semantic

type: E — exclusive or scalar, N — non-exclusive or set variable

AM — morphological analysis; AS — multilevel analysis;

TL — lexical transfer; TS — structural transfer;

GS — syntactical generation; GM — morphological generation

Variables RUS-FRA Continued

Level	Type	AM	AS	TL	TS	GS	GM	Names and values	Comments
A	N	*	*	*	*	*	*	P=(1, 2, 3)	PERSON: first, second, third
A	N	*	*	*	*			T=(PRE, PAS, IMP)	TENSE: present, past, imperative
C	N	*	*	*	*			TN=(PR, PN, NB)	NOUN TYPE: proper, pronoun, numeral
C	N	*	*	*	*			DGR=(COM, SUP, EGL)	DEGREE: comparative, superlative, equal
C	N	*	*					FM=(FOC, FOL)	ADJECTIVE FORM: short, long
A	N	*	*	*	*	*	*	G=(F, M, N)	GENDER: feminine, masculine, neuter
A	N	*	*	*	*			A=(P, I)	ASPECT: perfective, imperfective
A	N	*	*	*	*			N=(S, P)	NUMBER: singular, plural
A	N	*	*	*	*	*	*	RF=(R, N, PF)	VOICE: reflexive, active, passive
P _i	N	*						FN=(A, U, I, E, O, W)	(a) AMBIGUITY SIGN of the root with ending A-YA; U-YU; I-YI; E; O; Q-J or zero (b) IMPERATIVE TYPE: I for tonic, W for atonic
L _i	N	*						DICT=(1, 2, 3)	NUMBER of DICTIONARY for MORPHOLOGICAL ANALYSIS 1 for roots and prefixes beginning with A to O included; 2 for roots and prefixes beginning with P to Z 3 for endings; this variable does not belong to the mask and it is proper to AM
L	E	*	*	*	*	*	*	CS=(RLT, ORD, POS, IDF, DNG, QAL, ACT)	SEMANTIC CODE for adjective and abstract noun: relative-denominative adjective, ordinal, possessive, indefinite, denegative, qualitative adjective, action noun
V	N	*	*	*	*	*	*	CPI=(ACC, GEN, DAT, INS, INF, CTO, CTB, LIO, PRP, VID, ATR)	FIRST COMPLEMENT TYPE: direct object, genitive, dative, instrumental, infinitive sentence, subordinating sentence (with KHTO, KHTOBYI or indirect-interrogative), prepositional group, origin-direction of movement, attribute

Variables RUS-FRA Continued

Level	Type	AM	AS	TL	TS	GS	GM	Names and values	Comments
V	N	*	*	*	*			CP2=(GEN, DAT, INS, INF, CTO, CTB, PRP, VID, ATR)	SECOND COMPLEMENT TYPE: (idem)
V	N	*	*	*	*			ISJ=(IFS, CTS, CBS, GNG)	SUBJECT TYPE of IMPERSONAL SENTENCE: infinitive, subordinating sentence (with KHTO, KHTOBYI), noun group in genitive
V	N	*	*	*	*			PG=(2V, 2NA, 2OB, 2CR, 2POD, 2ZA, 4IZ, 4BEZ, 4DL, 4DO, 4OT, 4S, 4U, 5K, 5PO, 6V, 6NA, 6OB, 7MZ, 7ZA, 7NAD, 7PRD, 7POD, 7S)	TYPE OF THE PREPOSITION REQUIRED BY "PRP" IN CP1 or CP2: with 2=accusative — V, NA, O, KHEREZ, POD, ZA; with 4=genitive — IZ, BEZ, DLYA, DO, OT, S, U; with 5=dative — K, PO; with 6=prepositional — V, NA, O; with 7=instrumental — MEZHDU, ZA, NAD, PERED, POD, S
V	N	*	*	*	*			PGA=(idem)	TYPE OF THE PREPOSITION USED IN PREPOSITIONAL GROUP
V	N	*	*					ITV=(OTDO, OTK, SNA, IZV, IZNA, SV)	GOVERNED PREPOSITIONAL PAIR: OT ... DO; OT ... K; S ... NA; IZ ... V; IZ ... NA; S ... V
L	N	*	*					VL=(DM, LI, QU, IT)	MARK of DESIGNATIVE ELEMENT ABSORBED by a GROUP demonstrative-consecutive, indirect-interrogative relative, direct-interrogative
A	N	*	*	*	*			NG=(NE)	MORPHOLOGICAL NEGATION
S	N	*	*	*	*			CRL=(1, 2)	COORDINATING CONJUNCTION PAIR: 1-first; 2-second
L	N	*	*	*	*	*		ANM=(ANI, INA, MAS)	ANIMATION: animate entity, inanimate object, male with feminine flexion
V	N	*	*					NC=(2, 4, 7)	CASE for ADVERBIAL USING of NOUNS: accusative (VESQ GOD), genitive (PERVOGO MAYA), instrumental (NOKHQYU)

Variables RUS-FRA Continued

Level	Type	AM	AS	TL	TS	GS	GM	Names and values	Comments
V	N	*	*					AT=(1, 4, 7)	CASE for ATTRIBUTIVE USING of NOUNS: 1=nominative, 4=genitive, 7=instrumental
L	E	*	*	*	*			NAT=(TMP, LEU, DIR, PRV, CSQ, CSE, BUT, CDL, ACP, PAR, CON, MSE, RFC, COL, OBJ, ELT, GRP, UNT, SCN, IZM, TMS, RCM)	SEMANTIC CODE for NOUNS, ADVERBS and ADJECTIVES time, localization, direction, origin, consequence, cause, aim, condition, accompaniment, instrument, concession, measure, intensity of positive, colour, physical object, stuff, collective, measuring unit, science, other abstract noun, measure kind, intensity of comparative
V	N	*	*					SI=(1)	MARK VERBS governing infinitives whose subject is different
S	N	*	*					ESP=(ITC, AVG)	POTENTIALITY or REALITY of COMMA ABSORPTION introductory words or complex conjunctions
P	E	*						HMC=(A, OJ, U)	MARK of VARIANT ENDING if HOMOGRAPHY A — plural of nouns; OJ — masculine nominative of adjectives; U — genitive or prepositional of nouns
C	N	*	*	*	*			MD=(IFF, VRB, PRT, GRD, ADJ)	MOOD: infinitive, personal, participle, gerundive, adjective)
S	E	*	*	*	*			K1=(GN, GA, PH, CR, GB, MD, GP, AV)	SYNTAGMATIC CLASS: noun group, adjective group, sentence, comparative complement, numeral group, sentence without subject (infinitive, gerundive, participle), prepositional group, adverbial group
S	E	*	*	*				ANF=(ELD, ELP, RLT, PM)	ANAPHORIC or ELLIPTIC restitution of the word: verb, adjective, relative pronoun, other pronoun
F	E	*	*	*	*	*		FT=(PP, CB, CJ, AX, DT, CT, DEB, FIN, MF)	TERMINAL NODE SYNTACTICAL FUNCTION: preposition, subordinating, coordinating, conjunction, auxiliary verb, article and similar determinant, start and end of sentence, modifier

Variables RUS-FRA Continued

Level	Type	AM	AS	TL	TS	GS	GM	Names and values	Comments
F	E	*	*	*	*	*	*	LX=(GOV, PX, HM, EG, ED)	LEXICAL FUNCTION of the CONSTITUENT ELEMENTS IN EACH LEXICAL UNIT: head word, prefix, homograph, left qualifier, right qualifier
R	E	*	*	*	*	*	*	AG=(A1, A2, A3, A4, A5, A6, MS)	ARGUMENTS: agent, patient, beneficiary, patient-subject, coordinate element, relative sentence, measure complement
A	E	*	*	*	*	*	*	EMP=(THM, THN, RLF, NEG)	EMPHASIS and NEGATION: simple emphasis, emphasis + negation, soft emphasis, negation
A	E	*	*	*	*	*	*	DGA=(PLS, MNS, IDM)	DEGREE TYPE: superiority, inferiority, equality
F	N	*	*	*	*	*	*	ABS=(A1, A2, A3, SJ, AX, PP, SB)	MARK of the absorbed constituents: agent, patient, beneficiary, subject, auxiliary, prepositional group, subordinate clause
V	N	*	*	*	*	*	*	TVB=(TRS, SJA)	TACTICAL VERBAL VALENCY: possibility of government of the direct object, of the subject
S _t	N	*	*	*	*	*	*	REQ=(I)	TACTICAL MARK of rule application, i.e. copying results of the analysis from non-terminal group on its terminal head
S	N	*	*	*	*	*	*	MRQ=(CARD, RELAT, ADPX, QUAL, GENIT)	MARK of the absorbed constituents: numeral, relative clause, prefix, adjective, noun group in genitive by other noun group
C	E	*	*	*	*	*	*	KF=(NM, AQ, VB, IV)	FRENCH FLEXIONAL CLASS: noun, adjective, verb, other
C	E	*	*	*	*	*	*	DF=(ADV, CAR, ACT, AGT, ABL, BIL, DIR)	DERIVATIONS: adverb from adjective, noun from adjective, abstract noun from verb, agent noun or adjective from verb, potential passive adjective, idem noun, direct derivation
C	E	*	*	*	*	*	*	CDF=(ADS, ADMS, ADSF, RCT)	DIRECT SUBSTANTIVATION of ADJECTIVES: with gender from lang. source, masculine, feminine, nationality

Variables RUS-FRA Continued

Level	Type	AM	AS	TL	TS	GS	GM	Names and values	Comments
S	E		*	*	*			ART=(DEF, IDF, ABS)	REALIZED NOUN DETERMINATION: with definite, indefinite article, without article
A	E		*	*	*	*		TF=(PRE, IPF, FUT, PAS)	TENSE: present, imperfect, future, simple past
C	E		*	*	*	*		MF=(IND, SBJ, CDL, IMP, PRT, GRD, IFF)	MOOD: indicative, subjunctive, conditional, imperative, participle, gerundive, infinitive
A	E		*	*	*			AF=(1, 2)	COMPOUND TENSE: with one auxiliary, with two
P	E		*	*	*	*		SXF=(ION, URE, RIE, FMZ, NCE, ITE, ESS, EURF, ETE, IRG, PFM)	FEMININE NOUN DERIVATION: suffixes (AT)ION, (AT)URE, ERIE, E, (A, E)NCE, ITE, ESSE, EUR, ETE, irregular, participle form
P	E		*	*	*	*		SXM=(MENT, AGE, MSZ, PMS)	MASCULINE NOUN DERIVATION: suffixes EMENT, AGE, zero, participle form
P	E		*	*	*	*		AGN=(RIC, EUS, ANT)	AGENT DERIVATION: suffixes EUR-RICE; EUR-EUSE, ANT or ENT
A	E		*	*	*	*		NBR=(SIN, PLU)	FRENCH NUMBER: singular, plural
S	E		*	*	*			PRG=(GPEN, GPA, AJQ, GPD)	QUALIFIERS MUST BE CONSTRUCTED: with preposition EN, A, anyone, DE
V	E		*	*	*	*		ADF=(MON, TEL, BEL, ORD)	TYPE of ADJECTIVES to BE PUT before NOUNS: MON- takes place of the article; TEL — asks UN before; BEL — nothing; ORD — asks LE before
V	E		*	*	*			AVF=(JAM, TRES, TOT)	TYPE of ADVERBS CONCERNING their PLACE: JAM — verbal modifier but after auxiliary, TRES — adjective modifier; TOT — clause modif.
P	E		*	*				TOURN=(1)	LEXICAL TURN to verify in TS.
P	E		*	*				RSM=(0, 1, 2, 3, 4)	NODE POSITION in the SUB-TREE representing a LEXICAL TURN
S	E		*	*				LCL=(EN, AU, ASA)	FRENCH PREPOSITION for LOCALIZATION: EN, A=LE, A without article

Variables RUS-FRA Continued

Level	Type	AM	AS	TL	TS	GS	GM	Names and values	Comments
V	E		*	*	*			MFG=(SBJ, CDL, IFA, IFV)	GOVERNED MOOD: subjunctive, conditional, A+infinitive, infinitive without preposition
V	E		*	*				MPP=(ACC, IAC, IDE, NAQ, SKDE)	IRREGULAR TRANSFER of RUSSIAN PREPOSITIONS: any rus. prep.=direct object, instrumental=direct object, instrumental=DE, NA=DE, K=DE
P	N		*	*	*	*		TPN=SJA, ATN, TNQ, OBD)	FRENCH PRONOUN TYPE: atonic subject, atonic indirect object, tonic, atonic direct object
S _i	N		*	*				RFL=(RF1, RF2, RF3)	RUSSIAN REFLEXIVITY: 1 — passive present=potential, 2 — reflexive present participle=potential, 3 — all reflexive forms=passive
S	N		*	*	*			DIF=(MRE, MRN, POT, ACT, AGT)	DERIVATION DEFECTIVITY: adverb (aff. or negat) negative adverb, potential adjective, verbal noun, agent
V	N		*	*				AX=(ETRE, ETAT)	AUXILIARY: verb asks ETRE; it asks AVOIR but its past participle is active and not passive when used as adjective
A	N		*	*	*	*		NGF=(IN, PEU, NON)	FRENCH MORPHOLOGICAL NEGATION: by prefix IN, turn with PEU, PETIT; prefix NON-
V	N		*	*	*			MPC=(QUE, DE, A)	AUXILIARY ELEMENTS to joint to compound conjunctions or prepositions: QUE, DE, A
S _i	N				*			TRC=(RCH, RCB)	MARK of copying results from top or bottom nodes FRENCH ENDING VERBAL SYSTEMS
P	E					*		PV1=(PR1, PR2, PR4)	SINGULAR PRESENT IN: -e/-t/-zero
P	E					*		PV2=(IF1, IF2, IF3, IF4, IF5, IF7)	INFINITIVE: -er/-r/-ir/-re/-dre/-oir
P	E					*		PV3=(FT1, FT2, FT3, FT5)	FUTURE: -era/-ra/-ira/-dra

Variables RUS-FRA Continued

Level	Type	AM	AS	TL	TS	GS	GM	Names and values	Comments
P	E							* PV4=(PC1, PC2, PC3, PC4, PC5, PC6)	PAST PARTICIPLE: -é/zero, e /-i, -ie/-u, -ue/-t, -te/-s, -se
P	E							* PV5=(SV1)	PRESENT SUBJUNCTIVE and "ENT" in INDICATIVE
P	E							* PV6=(SV2)	IMPERFECT and PRESENT PLURAL 1-st, 2-d PERS.
P	E							* PV7=(PS1, PS2, PS3, PS4)	SIMPLE PAST: -a/-t/-it/-ut
P	E							* ALT=(ALT1, ALT3)	ROOT ALTERNATION: 1 — jeter, appeler, broyer; 2 — finir, paraître
P	E							* DN=(DN1, DN2, DN3, DN4, DN5, DN6, DN7)	FRENCH ENDING NOMINAL SYSTEMS: /-, s, e, es/; /e, es, e, es/; /-, x, e, es/; /x, x, e, es/; /s, s, e, es/; /eur, eurs, euse- rice-eresse/; /-, s/
P	E							* DD=(DD1, DD2, DD3, DD4, DD5, DD6, DD7)	FLEXIONAL BEHAVIOUR CODE: 1- normal; 2-invariable form; 3-form containing the suffix; 4-elidable form; 5-prefix or punctuation to bind on the left; 6-postfix or punctuation to bind on the right; 7- hyphen
P	N							* IFX=(AT, QC, IBL)	AUXILIARY ALTERATION CODE: AT=add AT to suffixes ION, EUR, URE; QC=transform QU, GU in C, G before A; IBL=potential adjective form is in IBLE and not ABLE
P	N							* ALG=(ALG1, ALG2)	SPECIAL CODE for VERBS ENDING in ... GER in order to transform them before endings in O, A, U; i.e. MANG- EONS

Notes

1. For this implementation, we choose not to generate executable machine code in order to save space.
2. We will use either 'grammar' or 'TG', 'rule' or 'TR', 'transformational system' or 'TS'.
3. Note that 'governor' is the translation of *régisseur* in French. 'Head-' translates *gouverneur*.

References

- Axtmeyer, M., Gaudey, J. and Torre, L. (1979). *Indexage morpho-syntaxique du russe*. Grenoble: Rapport de recherche G.E.T.A.
- Boitet, Ch. (1976). Un essai de réponse à quelques questions théoriques et pratiques liées à la traduction automatique. Définition d'un système prototype. Thèse d'Etat. Grenoble: Université Scientifique et Médicale de Grenoble.
- (1976). Problèmes actuels en traduction automatique: un essai de réponse. In *Proceedings of COLING 76*. Ottawa: University of Ottawa, Department of Linguistics.
- (1977). Mechanical translation and the problem of understanding natural languages. In *Proceedings of Colloque Franco-Soviétique*. Moscou: Université de Moscou, Laboratoire de Calcul.
- (1979). *Extension de ROBRA et utilisation de l'environnement de contrôle interactif à l'exécution*. Grenoble: Rapport de recherche G.E.T.A.
- (1979). Automatic production of CF and CS analyzers using a general tree-transducer. In *Proceedings of 2. Internationales Kolloquium über Maschinelle Uebersetzung, Lexikographie und Analyse*. Saarbrücken, Grenoble: R. R. No. 218, Université Scientifique et Médicale de Grenoble, Institut de Mathématiques Appliquées et Informatique.
- Boitet, Ch., Guillaume, P. and Quézel-Ambrunaz, M. (1978). Manipulation d'arborescences et parallélisme: le système ROBRA. In *Proceedings of COLING 78*. Bergen: University of Bergen, Nordic Institute, Department of Computational Text Processing.
- Chauché, J. (1974). Transducteurs et arborescences. Etudes et réalisations de systèmes appliquées aux grammaires transformationnelles. Thèse d'Etat. Grenoble: Université Scientifique et Médicale de Grenoble.
- (1975). *Présentation du système C.E.T.A.* Grenoble: Rapport de recherche G.E.T.A. No. G.3100.A.
- Chauché, J., Guillaume, P. and Quézel-Ambrunaz, M. (1972). *Le système A.T.E.F. (Analyse de Textes en Etats-Finis)*. Grenoble: Rapport de recherche G.E.T.A. No. G.2600. A.
- Communication Groupée. (1977). *Où en est le G.E.T.A. début 1977?* Troisième Congrès Européen sur les Systèmes et Réseaux Documentaires, Franchir la Barrière Linguistique, Luxembourg, 1977. Munich: Verlag Dokumentation.
- Guillaume, P. (1978). *Le modèle de transformations d'arbres: ROBRA*. (Présentation des caractéristiques externes de la grammaire complétant celles du modèle C.E.T.A.). Grenoble: Rapport de Recherche G.E.T.A. No. G.3200.A.
- Jaeger, D. (1978). *SYGMOR. (Système de Génération MORphologique)*. Grenoble: Rapport de Recherche G.E.T.A. No. G.3300.A.
- (1979). *SYGMOR. Nouvelle version. Exécuteur en Assembleur 360*. Grenoble: Rapport de Recherche G.E.T.A.
- Nedobejkine, N. (1976). Niveaux d'interprétation dans une traduction multilingue: application à l'analyse du russe. In *Proceedings of COLING 76*. Ottawa: University of Ottawa, Department of Linguistics.
- (1973). Application du système A.T.E.F. à l'analyse morphologique de textes russes. In *Proceedings of COLING 73*. Pisa, Firenze: Leo S. Olschki (1977).
- Quézel-Ambrunaz, M. (1978). *ARIANE. Système interactif pour la traduction automatique multilingue. (Version II)*. Grenoble: Rapport de Recherche G.E.T.A. No. G.3400.A.
- (1979). *Transfert en ARIANE-78. Le modèle TRANSF*. Grenoble: Rapport de Recherche G.E.T.A.
- Quézel-Ambrunaz, M. and Guillaume, P. (1973). Analyse automatique de textes par un système d'états-finis. In *Proceedings of COLING 73*. Pisa, Firenze: Leo S. Olschki (1977).

- Thouin, B. (1976). Système informatique pour la génération morphologique de langues naturelles en états-finis. In *Proceedings of COLING 76*. Ottawa: University of Ottawa, Department of Linguistics.
- Vauquois, B. (1975). *La traduction automatique à Grenoble*. Documents de Linguistique Quantitative No. 24. Paris: Dunod.
- (1976). Automatic Translation. A Survey of Different Approaches. In *Proceedings of COLING 76*. Ottawa: University of Ottawa, Department of Linguistics.
- (1977). L'évolution des logiciels et des modèles linguistiques pour la traduction automatique. In *Proceedings of Colloque Franco-Soviétique*. Moscou: Université de Moscou, Laboratoire de Calcul.