

Internal distribution code:

- (A) [] Publication in OJ
(B) [X] To Chairmen and Members
(C) [] To Chairmen
(D) [] No distribution

D E C I S I O N
of 9 July 2002

Case Number: T 1177/97 - 3.5.1
Application Number: 87400004.5
Publication Number: 0274281
IPC: G06F 15/38
Language of the proceedings: EN

Title of invention:

Method using a programmed digital computer system for translation between natural languages

Patentee:

SYSTRAN S.A.

Opponents:

Siemens Nixdorf Informationssysteme Aktiengesellschaft
Logos Computer Integrated Translation GmbH

Headword:

Translating natural languages/SYSTRAN

Relevant legal provisions:

EPC Art. 52(1)-(3), 56, 104(1)

Keyword:

"Inventive step (no)"
"Problem-and-solution approach: treatment of non-technical aspects"
"Admissibility of late-filed documents (no)"
"Different apportionment of costs (no)"

Decisions cited:

T 0052/85, T 0163/85, T 0769/92, T 1194/97, T 0641/00

Catchword:

1. The use of a piece of information in a technical system, or its usability for this purpose, may convey a technical character to the information itself in that it reflects the properties of the technical system, for instance by being specifically formatted and/or processed. Such information when used in or processed by the technical system may be part of a technical solution to a technical problem and form the basis for a technical contribution of the invention to the prior art.

2. Information and methods related to linguistics may thus in principle assume technical character if they are used in a computer system and form part of a technical problem solution.



Case Number: T 1177/97 - 3.5.1

D E C I S I O N
of the Technical Board of Appeal 3.5.1
of 9 July 2002

Appellant: SYSTRAN S.A.
(Proprietor of the patent) 26 bis, avenue de Paris
F-95230 Soisy-sous-Montmorency (FR)

Representative: Derambure, Christian
Bouju Derambure Bugnion
52, rue de Monceau
F-75008 Paris (FR)

Respondents: Siemens Nixdorf Informationssysteme
(Opponent 02) Aktiengesellschaft
Fürstenallee 7
D-33102 Paderborn (DE)

Representative: Epping Hermann & Fischer
Ridlerstrasse 55
D-80339 München (DE)

(Opponent 03) Logos Computer Intregrated Translation GmbH
Mergenthaler Allee 79-81
D-65760 Eschborn (DE)

Representative: Nix, Frank Arnold, Dr.
Kröckelbergstrasse 15
D-65193 Wiesbaden (DE)

Decision under appeal: Decision of the Opposition Division of the
European Patent Office posted 7 October 1997
revoking European patent No. 0274281 pursuant
to Article 102(1) EPC.

Composition of the Board:

Chairman: S. V. Steinbrener
Members: R. R. K. Zimmermann
S. C. Perryman

Summary of Facts and Submissions

I. European patent number 0 274 281 was granted to the appellant with effect of 10 June 1992 on the basis of a European patent application filed in 1987.

II. The invention to which the patent relates is in the field of machine translation of natural languages and concerns the so-called SYSTRAN translation system, the development of which, with leading contributions from the appellant, goes back to the 1960's.

III. Oppositions were filed by respondents O2 and O3 against the patent in its entirety on 9 and 10 March 1993, respectively, *inter alia* on the grounds of Article 100(a) EPC in respect to non-patentable subject-matter under Article 52(2)(c) EPC and to lack of novelty and inventive step. The prior art cited includes among others the following documents:

D4: M. Thiel "Wörterbuchsuche" in: "Automatische Lemmatisierung, Zielsetzung und Arbeitsweise eines linguistischen Identifikationsverfahrens", 3. Berichtsteil, Linguistische Arbeiten 15, Sonderforschungsbereich Elektronische Sprachforschung, Universität des Saarlandes, Saarbrücken, 1976

D18: Peter P. Toma et al. "Optimization of SYSTRAN System", RADC-TR-72-73 Final Technical Report submitted by LATSEC, Inc., 1972, Rome Air Development Center, Air Force Systems Command, Griffiss Air Force Base, New York

The opposition division revoked the patent for lack of inventive step with a decision posted on 7 October 1997.

IV. The appellant lodged an appeal against the decision, filing the notice of appeal and paying the appeal fee on 5 December 1997. The written statement setting out the grounds of appeal was filed on 17 February 1998.

V. At oral proceedings which took place on 9 July 2002 in the presence of the appellant and respondent 03, the appellant replaced all previous versions of the claims submitted by following claim 1:

"1. A method for translation between source and target natural languages using a programmable digital computer system, the steps comprising:

(a) storing in a main memory of the computer system a source text to be translated;

(b) scanning and comparing such stored source words with dictionaries of source language words stored in a memory and for each source text word for which a match is found, storing in a file in main memory each word, and in association with each such word, coded information derived from such dictionary for use in translation of such word, the coded information including memory offset address linkages to a memory in the computer system where grammar and target language translations for the word are stored;

(c) analysing the source text words in its file of words, a complete sentence at a time, and converting the same into a sentence in the target language utilizing the coded information and including the steps of

(1) utilizing the memory offset address linkages for obtaining the target language translations of words from a memory; and

(2) reordering the target language translation into the proper target language sequence;

the steps of analysing additionally comprising the steps of analysing each source word in multiple passes through each sentence of the source text, assigning codes thereto, considering all the codes which previous passes have attached to a word and assigning target language synthesis codes attached to the meaning with which the code functions in the sentence, placing the word into a form corresponding to the target language dependent upon the analysis and consideration of all relevant codes assigned to the words, wherein said dictionaries of source language words comprise entries containing a source language stem, the coded information associated to such entry comprising an offset address linkage relating to the set of valid endings permitted for said source language stem, and

said method further includes the steps of:

- storing a dictionary of high frequency source words and associated offset address linkages, the offset address linkages identifying the storage location of grammar and meaning information for the source words;
- comparing each source language text word with the high frequency dictionary words and, upon detecting an equality with a word, storing the word and associated offset address linkages together in a high frequency file; and
- if no equality is detected, storing the word in a low frequency file, and

for each source text word in the low frequency file, the step of comparing such word with dictionaries of source language words comprises the steps of:

- inspecting said dictionaries to determine whether a particular entry thereof matches said source text word,
- if no match is found, dropping the last letter of said source text word and;
- repeating the sequence of said steps of inspecting and dropping the last letter until a match is found with a source language stem entry of said dictionaries, the number of letters dropped being less than a predetermined number representing the maximum ending length for said source text word, and upon finding a match
- inspecting the set of valid endings attached to said source language stem entry until finding a match between said source text word and a stem and ending combination,

wherein:

in the case where said step of inspecting reveals that a particular stem entry matches a source text word having at least one letter dropped, the chopped-off ending of said source text word, made of the sequence of dropped letters, is compared to the set of valid endings permitted for said stem entry and, upon detecting an equality in this comparison, grammar information is stored such as the gender, number, declensional case or conjugational tense corresponding to the identified valid ending, and

second and succeeding idiom words, if any, in a source idiom are stored in at least one dictionary in memory, the high frequency dictionary includes the first words of source language idioms and associated therewith address linkages to second and subsequent words in the same idiom which are located in the same idiom dictionary, the first word and subsequent words, if any, of an idiom having an associated target language meaning stored in association therewith; and during comparison with the high frequency dictionary comprising the steps of:

- (a) detecting those words that are equal and are first words of idioms ;
- (b) utilizing the offset address linkages to locate the additional word or words in the same idiom located in the idiom dictionary ;
- (c) comparing the located further words in the idiom with the words in the source text which follow the first idiom word for an equality and ;
- (d) upon detecting such an equality, storing the first idiom word together with the target language meaning into the high frequency file, and

a separate file is stored containing limited semantic numbers for each principal word, the limited semantic numbers indicating all participating words, the limited semantic numbers of participating words being attaches [sic] to the principal words in the same sequence in which the participating words form compound expressions with the principal word, and

during the steps of analysing and searching the file of words to establish whether a principal word is surrounded by supplementary words, if supplemental

words are established, the limited semantic numbers of the supplementary words are compared against the limited semantic numbers stored in the limited semantic dictionary for each principal word, if a match is encountered, translating the principal and supplementary words forming a compound into the corresponding meaning."

- VI. At the oral proceedings, respondent 03 sought to introduce a new prior art document on dictionary look-up techniques and, furthermore, a post-published United States patent, both documents authored by the designated inventor of the present invention. The appellant objected to the late filing of these documents and requested that the new prior art document should not be admitted into the proceedings, or if it was admitted, that the case be remitted to the first instance for further examination. After considering whether exceptional circumstances justified the late filing, the Board decided not to admit the document to the proceedings.

At the end of the oral proceedings the Board's decision was announced.

- VII. Regarding the question of inventive step the appellant referred to document D18 from which the closest prior art was derivable. Document D18, however, did not disclose the longest match principle for use in a low frequency dictionary look-up process. Because of this difference the invention provided a very efficient search strategy, in terms of memory requirements and processing time, for locating the source text words in the low frequency file of the system. Using the longest

match principle as a key element of the search algorithm was a remarkable improvement of older SYSTRAN versions.

The longest match principle might have been applied before to dictionary search, but it was not used before with SYSTRAN-type translation systems.

Although document D4 referred to a longest-matching method ("longest-matching Verfahren") it mentioned this only as a theoretical possibility for determining flexion endings, but it did not give any hints to use such a method in order to look for the translation of the global word, i.e. stem plus ending, in a stem dictionary. In addition, an alternative "shortest matching" was mentioned as the theoretically better alternative since shorter endings occurred more frequently than longer endings. Moreover, there was no suggestion to use the longest-matching method in only a part of the translation method, namely only in the SYSTRAN low-frequency file for determining the appropriate stem and ending among the set of valid endings. Document D4, actually, concerned the translation of a German language text, but it did not teach how to translate other languages having a greater variety of syntactical modes. The claimed invention was thus not rendered obvious by document D4, nor by any other combination of prior art documents cited against the patent.

VIII. Accordingly, the appellant requested that the decision under appeal be set aside and that the patent be maintained on the basis of the set of claims filed as main request at the oral proceedings on 9 July 2002.

Referring to possible claim deficiencies the appellant expressed its willingness to amend the claims when the main issues regarding patentability had been settled.

Both respondents O2 and O3 - either orally or in writing - requested that the appeal be dismissed.

IX. Although there was no dispute about document D18 as being the closest prior art, the respondents disagreed with the proposition that the longest-matching method provided an inventive contribution to the prior art. These sort of techniques were fundamental to natural language processing; they were employed in a great many of natural language processing systems. If all features and effects of these techniques had not been explicitly mentioned in a single document it was simply because they were implicit to the skilled person and explicit reference was unnecessary. In particular, as may be seen from document D4 the longest-matching method was a common option well-known in the field of automatic translation systems for searching words, stems or endings in dictionaries. This type of search strategy was also disclosed in other prior art documents cited against the patent.

X. Respondent O3 observed that the appellant, although having requested oral proceedings entailing considerable costs to the parties to the appeal proceedings, had not submitted any new arguments or facts, but on the contrary, various claim deficiencies already removed in the first instance proceedings had been reintroduced into the claims. This was an unreasonable behaviour justifying compensation.

XI. Respondent O3, for these reasons, asked for an apportionment of costs incurred by the oral proceedings of 9 July 2002 in its favour.

The appellant disagreed and requested that the request for an apportionment of costs should be refused.

Reasons for the Decision

1. The appeal complies with the requirements of Articles 106 to 108 and Rules 1(1) and 64 EPC and is thus admissible.

The appeal, however, is not allowable since the subject-matter of claim 1 of the appellant's single request does not comply with the requirement of inventive step as set out in Articles 52(1) and 56 EPC.

Inventive step

2. According to Article 56 EPC, an invention shall be considered to involve an inventive step, if having regard to the state of the art, it is not obvious to a person skilled in the art. This legal definition of inventive step is usually applied by using the so-called problem and solution approach, which requires analysis of the invention in terms of a technical solution to a technical problem (see the fourth edition of the "Case Law of the Boards of Appeal of the European Patent Office", 2002, European Patent Office, pages 101 f.).

3. Since the solution as well as the problem solved by an invention should be of a technical nature the problem and solution approach may raise questions when the invention comprises non-technical aspects or elements. Such difficulties are to be resolved by taking due care to define the technical field to which the invention belongs, the scope of technical expertise and skills expected to be applied by the technical person in this particular technical field, and the correct formulation of the technical problem actually solved. Features of the invention which do not form part of the technical solution to the technical problem have to be disregarded in the assessment of inventive step (see for example decision T 641/00 Two identities / COMVIK, to be pub. in OJ EPO).

Present claim 1 is directed to a method for translation between natural languages; accordingly it uses various linguistic terms and involves linguistic aspects of the translation process. This kind of subject-matter renders it necessary to construe the claim to determine the technical features of the method which alone are relevant to inventive step. It raises the even more basic question whether such linguistic concepts and methods may form part of a technical invention at all. The case law of the EPO provides various examples that even the automation of such methods does not make good a lack of technical character. So for example in T 52/85 (not pub. in OJ EPO), point 5 of the reasons, the mere use of a conventional general purpose computer for solving a problem in the field of linguistics and displaying information about semantically related expressions was found not to qualify as a technical contribution to the art.

On the other hand, coded information has been considered, on a case-by-case basis, as a patentable entity: in decision T 163/85 - Colour television signal / BBC, OJ EPO 1990, 379 a claim to a television signal was allowed since the signal was claimed in terms which inherently comprised the technical features of the television system in which it occurred. This decision was confirmed by a different board in T 1194/97 - Data structure product / PHILIPS, OJ EPO 2000, 525 and analogously applied to a record carrier characterised by a functional data structure of picture line synchronisations, line numbers and addresses. As a final example decision T 769/92 - General purpose management system / SOHEI, OJ EPO 1995, 525 may be cited in which a "transfer slip" providing a unitary slip format was not considered to be a presentation of information as such, but to be a user interface allowing the combination of two different management systems by a common input device and thus requiring technical considerations from the skilled person.

Hence, in accordance with this jurisprudence it seems to be common ground that the use of a piece of information in a technical system, or its usability for this purpose, may confer a technical character on the information itself in that it reflects the properties of the technical system, for instance by being specifically formatted and/or processed. Such information when used in or processed by the technical system may be part of a technical solution to a technical problem and form the basis for a technical contribution of the invention to the prior art.

In so far as the technical character is concerned it should be irrelevant that the piece of information is used or processed by a conventional computer, or any other conventional information processing apparatus, since the circumstance that such an apparatus had become a conventional article for everyday use does not deprive it of its technical character just as a hammer must still be regarded as a technical tool even though its use has been known for millennia. It would also be irrelevant that the invention involves semantic aspects of the information, or any "cognitive information content", since adding a non-technical component, or features, does not reduce a technical component of an invention to zero (see point 3.6 of the reasons and headnote 2 of the SOHEI-decision cited above).

The Board thus comes to the conclusion that information and methods related to linguistics may in principle assume technical character if they are used in a computer system and form part of a technical problem solution.

Implementing a function on a computer system always involves, at least implicitly, technical considerations and means in substance that the functionality of a technical system is increased. The implementation of the information and methods related to linguistics as a computerized translation process similarly requires technical considerations and thus provides a technical aspect to *per se* non-technical things such as dictionaries, word matching or to translating compound expressions into a corresponding meaning. Features or aspects of the method which reflect only peculiarities

of the field of linguistics, however, must be ignored in assessing inventive step.

4. By the end of the oral proceedings before the Board it was accepted by all the parties that document D18 dealing with an optimization of a prior art SYSTRAN system was the most relevant piece of prior art. As agreed by both parties at the oral proceedings, this document anticipates the claimed method in general, including a low frequency dictionary look-up process for translating source text words stored in a low frequency file (see in particular pages 8 to 14).

5. Neither this document nor any other of the cited prior art documents, however, disclose a low frequency dictionary look-up process on the basis of the so-called "longest match principle", meaning a look-up process which produces the dictionary entry with the longest stem matching the source text word. According to claim 1, the look-up process searches in the dictionaries for a source language stem entry matching the word, and, failing a match, drops the last letter from the word and again searches. This process of dropping the last letter and searching is repeated, subject to a predetermined limit, until a match between said source text word and a stem and ending combination is found, taking account of the valid endings allowed for this stem.

6. According to the patent in suit, the object of the claimed invention "is to improve the method ... so that it may be implemented more efficiently, particularly for translations between natural languages having a great variety of declensional and conjugational modes" (see page 3, lines 15 to 17).

Document D4 relating to automated search in dictionaries directly points to a character-by-character process, such as the claimed one, cutting off characters beginning from the end of the word until a valid stem and ending combination has been found, for performing longest matching (see page 43, last paragraph to page 44, first paragraph). It also mentions an alternative, the "shortest-matching method", compares the advantages and disadvantages of both alternatives and gives linguistic reasons why the "longest match principle" is to be preferred (at least for the German language). In particular, for the requirement of efficiency, it should be determined as early as possible that no longer ending is possible.

Although D4 specifically deals with the German language, it appears to the Board that on the basis of the rather general discussion in this document the use of the preferred alternative in the context of the above object is straightforward for a person having linguistic knowledge and would hence indeed be obvious, as was argued by respondent 03 at the oral proceedings.

7. Moreover, strictly speaking, it is even not apparent that the appropriate selection of the matching procedure contributes to the technical character of the invention as can already be seen from the argument

given in the preceding paragraph. Whereas in the Board's opinion the use of a low frequency dictionary and an algorithm of sequentially dropping letters as part of the matching process may be regarded as the result of a specific adaptation of the translation process for computer implementation and thus in principle as technical components of the claimed invention, the application of the "longest match principle" is in substance based on linguistic considerations as it is the natural language to be translated which determines whether the one or the other matching principle delivers better results. From a technical point of view, both alternatives are equivalent in that the respective different truncating steps must be straightforwardly translated into corresponding computer routines.

Applying the principles laid down by the Board in its COMVIK-decision cited above (see headnote II), the decision for one or the other matching principle does not seem to solve any technical problem and hence does not fall within the responsibility of a technically skilled person. It is rather a non-technical constraint determined by the linguistic expert and given to the skilled person as part of the framework of his task, namely implementing the known low frequency dictionary look-up process by applying the "longest match principle".

Choosing to apply the one or the other principle has clearly consequences for the technical implementation of the translation process since the computer routines have to work differently and the automated translation process will produce objectively different results,

technical differences which establish novelty. These technical differences, nevertheless, are not inventive since they originate from a non-technical constraint to the technical problem, the implementation of which is obvious.

8. It follows that the method of claim 1 lacks inventive step (Article 56 EPC) and hence is not patentable under Article 52(1) EPC.

Late filed documents

9. The respondent submitted an additional prior art document relating to "General Analysis Technique - The Dictionary Look-up" for the first time at the oral proceedings of 9 July 2002. The Board followed the appellant's request not to admit this document into the proceedings since it gave rise to new issues in the context of a rather complex subject-matter, which could not be expected to be dealt with at short notice.
10. The post-published patent, finally, lacks any relevance to the issue of obviousness over the prior art.

Apportionment of costs

11. Regarding the respondent's request for apportionment of costs incurred by the oral proceedings in its favour, the Board observes that according to Article 116 EPC it is a basic procedural right of parties to the proceedings in examination, opposition and appeal to be heard in oral proceedings before the responsible department. A decision deviating from the rule that each party to the proceedings has to meet the costs it

has incurred (Article 104(1) EPC) is reserved for very exceptional cases where reasons of equity render a different apportionment of costs necessary.

In the present case, the appellant defended its patent exercising its ordinary rights, i.e. arguing its case in the hearing and filing amended claims. It would need quite exceptional circumstances for the Board to be persuaded that someone was not in good faith exercising his legal rights but intending merely to cause other parties to incur costs. No such exceptional circumstances appear here. The amendments objected to by the respondent 03 were not the reason for needing oral proceedings. They can only be considered a minor irritation not justifying any apportionment of costs.

The respondent's request for an apportionment of costs in its favour is thus refused.

Order

For these reasons it is decided that:

1. The paper on General Analysis Technique submitted at the oral proceedings on 9 July 2002 is not admitted into the proceedings.
2. The appeal is dismissed.
3. The request for an apportionment of costs is refused.

The Registrar:

The Chairman:

M. Kiehl

S. V. Steinbrener