[From: *Tools for the trade: Translating and the Computer 5*. Proceedings of a conference ... 10-11 November 1983, ed. Veronica Lawson (London: Aslib, 1985)]

New developments in TITUS 4

A. A. Streiff

Information and Data Processing Department, Institut Textile de France, Boulogne-Billancourt, France

The TITUS 4 system was originally designed to produce abstracts in the form of sentences or phrases written in controlled syntax. It is now being improved, partly to give the user more flexibility in writing sentences, and partly so that the system can be implemented in other fields than abstracting services. Improvements being introduced to enhance TITUS 4's versatility include multiple-clause sentences. Certain restrictions, however, remain owing to linguistic problems associated with translation from one language to another.

INTRODUCTION

The TITUS 4 system was designed originally to receive abstracts in the form of sentences or phrases specially encoded for input to a computer in any one of four languages (English, French, German and Spanish).

The computer stores this input in suitable form and, on command, outputs the abstracts as sentences or phrases in one or more of the four languages specified in the command.

The TITUS 4 system enables abstracts to be written in sentences or phrases which closely resemble free-language sentences or phrases. Certain limitations or restrictions are, however, imposed by the computer program and by linguistic problems associated with translation from one language to another.

CONTROLLED SYNTAX LANGUAGE

The controlled syntax language used by the TITUS 4 system comprises two basic elements (Figure 1).

1. A sub-set of the whole vocabulary of a language containing all terms specific to a given field and a part of the basic vocabulary which is common to all fields.

2. A sub-set of all syntactic rules of a language. Though few in number, the syntactic rules of the TITUS 4 system allow the user to formulate a large variety of ideas.

GENERAL FLOW DIAGRAM OF TITUS 4 (Figure 2)

Document records to be introduced in the computer-held database must be encoded according to the specific TITUS 4 recording rules. All sentences of a record should match the syntactic rules recognised by the system and only certain terms picked out of the TITUS 4 pre-established vocabulary.

Records, whether written in English, French, German or Spanish, are interactively introduced into the computer by means of screen-type terminals.

Right use of recording rules is controlled, and lexical and syntactic validity is checked by 'generative grammars'. Any error or ambiguity arising during the input phase is detected by the program and causes the display of an error message or a question (polysemy, homography) requiring a statement from the user.

Each sentence is transformed into 'swivel language' before its storage. It is a binary form language whose particular structure permits very quick translation of a sentence in one or more of the four languages. After a sentence has been validated, 'generative grammars' index it automatically by detecting the descriptors contained in it. During the output phase, any document record stored in 'swivel language' is processed by 'output transformational grammars' which effect automatic translation.

TITUS 4 VOCABULARY

Sentences and phrases must consist only of authorised words. The authorised lexical elements are either 'fixed elements' or 'variable elements'.

1. Fixed elements

The fixed elements included in the program comprise:

- determiners (articles, demonstratives, possessives, quantifiers, cardinals...)
- conjunctions
- prepositions
- adverbs of degree
- negatives
- auxiliary verbs and modals.

2. Variable elements

The variable elements listed alphabetically in the lexicon comprise nouns, adjectives, verbs and adverbs.

2.1. Nouns

These may be single-word concepts or pre-coordinated multiple-word concepts. No lexical element can exceed 48 characters in length, including spaces.

2.2. Adjectives

These may be of three types:

(i) simple adjectives, which may be single-word or multiple-word adjectives
(ii) adjectives with complementation
(iii) past-participles formed from verbs in the lexicon, used adjectivally.

2.3. Verbs

The verbs in the lexicon are listed in their infinitive form. Most of the tenses and verbal elements (participles) are derived from the infinitive.

2.4. Adverbs

Common adverbs of frequency and manner are included in the lexicon.

LATEST IMPROVEMENTS IN TITUS 4

The main development in TITUS 4 since its implementation in 1980 is the introduction of subclauses (subordinate clauses, relative clauses and 'that' clauses) along with pronouns, giving the system more flexibility in writing sentences.

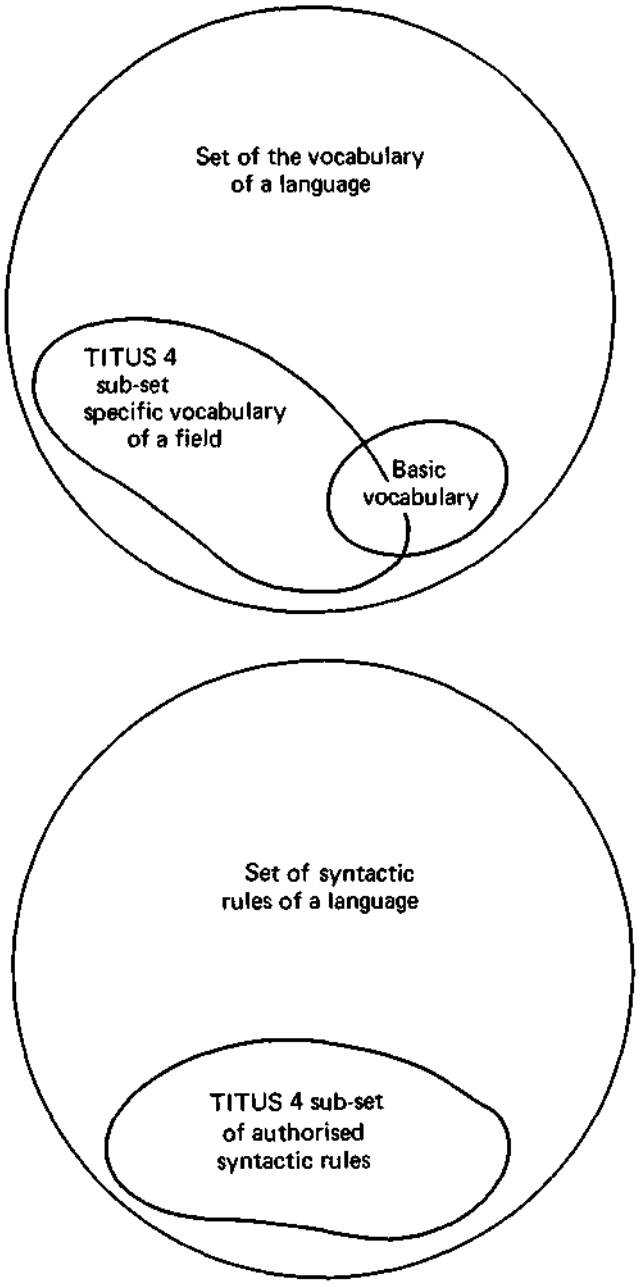

Figure 1. The controlled syntax language used by TITUS 4

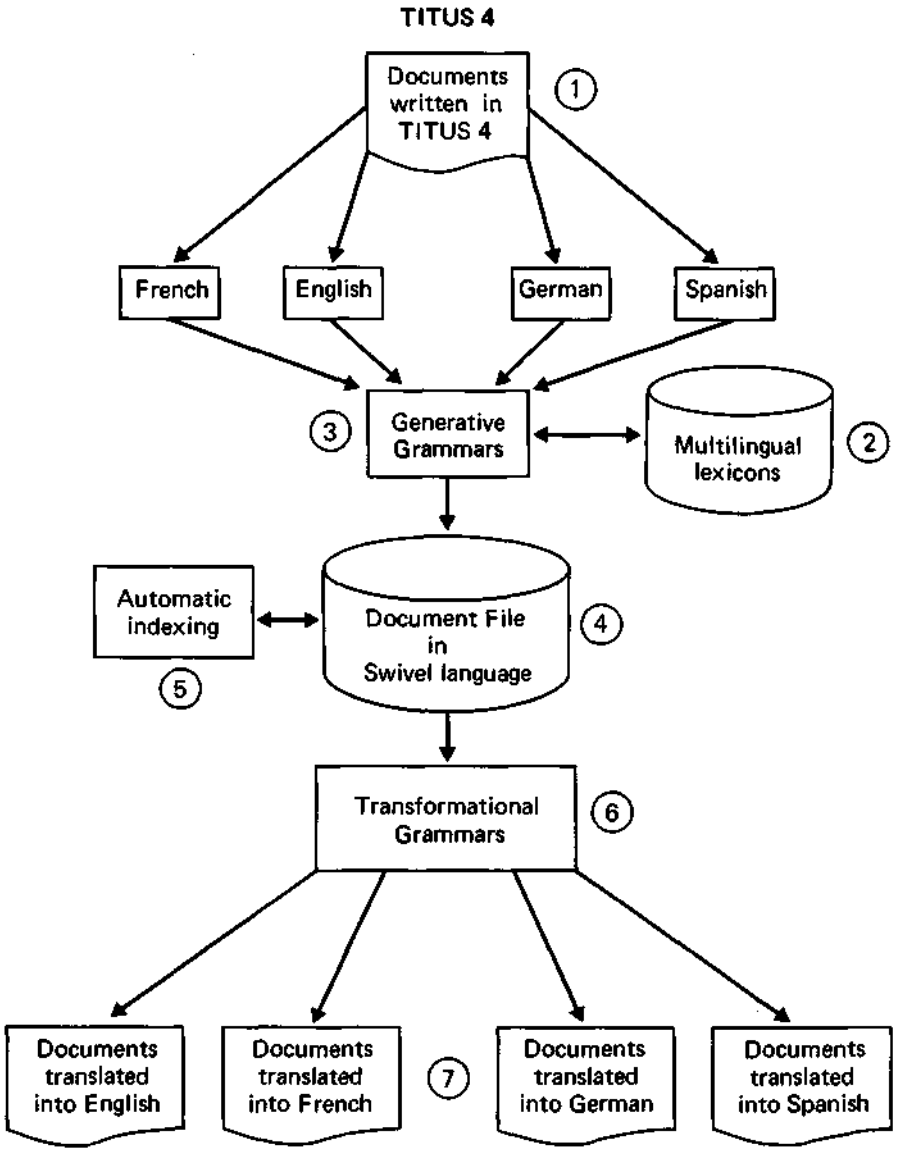

Figure 2. General flow diagram of TITUS 4

The basic pattern of a sentence or a clause (Figure 3) as established for TITUS 4, remains unchanged and the same writing rule will apply to any clause of the whole sentence.

A sentence may contain up to four clauses, the first being always the main clause.

1. Subordinate clauses

Subordinate clauses are introduced by subordinating conjunctions (although, if, when, since, because...).

2. 'That' clauses

Only those clauses which are the object of a verb in another clause are taken into account by the system.

example: The author states that this method may be improved.

3. Relative clauses

Relative clauses are introduced by the relative pronouns (who, which and that). These clauses may or may not be included in another clause (main clause or other subclause).

To avoid any confusion between some subordinating conjunctions (before, after, since, etc.) and prepositions, diacritical marks are used to indicate to the program the beginning of a subordinate clause. Relative pronouns and personal pronouns as well are considered as fixed elements and included in the programs. Because the system is interactive, antecedents of pronouns (either relative or personal) can be determined in the same way as is done for prepositions, by the use of diacritical marks.

A question/answer procedure is being set up to determine the antecedent of a pronoun whose antecedent is in a preceding clause. Another innovation in the near future will be the use of the imperative form of verbs, this form being generally used in command languages of users' manuals and similar documents.

RESTRICTIONS AND ADVANTAGES

Any user of the TITUS 4 system must comply with the writing rules imposed by the controlled syntax, and the training time to become familiar with it has been estimated at five to six days (full-time).

It will be easily understood that, since the TITUS 4 programs are interactive (question/answer procedure), the input of sentences or texts should be done by people who

have themselves written them according to the controlled syntax and the corresponding vocabulary.

One of the advantages frequently cited is the absence of cumbersome phrases or verbosity, thus much improving the clarity of the texts.

Any document or text entered can be interactively corrected after having been checked by the user himself from the four-language control listing edited after input has been completed.

CONCLUSION

We do not yet possess sufficiently sophisticated computers endowed with artificial intelligence (if indeed we ever shall) capable of interpreting all the intellectual subtleties and shades implicit in written expression in different languages.

For this reason TITUS 4 was based on a veritable 'controlled-syntax language' which can be processed by computer, but which is a simplified and formalised synthesis of natural language.

We hope TITUS 4 will not be restricted to abstracting and information services publishing multilingual bibliographic periodicals. The TITUS 4 system may be a useful and reliable tool for export-market-oriented industry, one of whose primary needs is to publish multilingual technical brochures, leaflets and notices and whose translation costs are nowadays of paramount importance. Another application of TITUS 4 could be its implementation in more sophisticated systems which automatically scan free language texts and pick out all the sentences matching the models defined by the controlled syntax.

AUTHOR

A.A. Streiff, Information & Data Processing Department,
Institut Textile de France, 35 rue des Abondances,
BP 79, 92106 Boulogne-Billancourt, France.

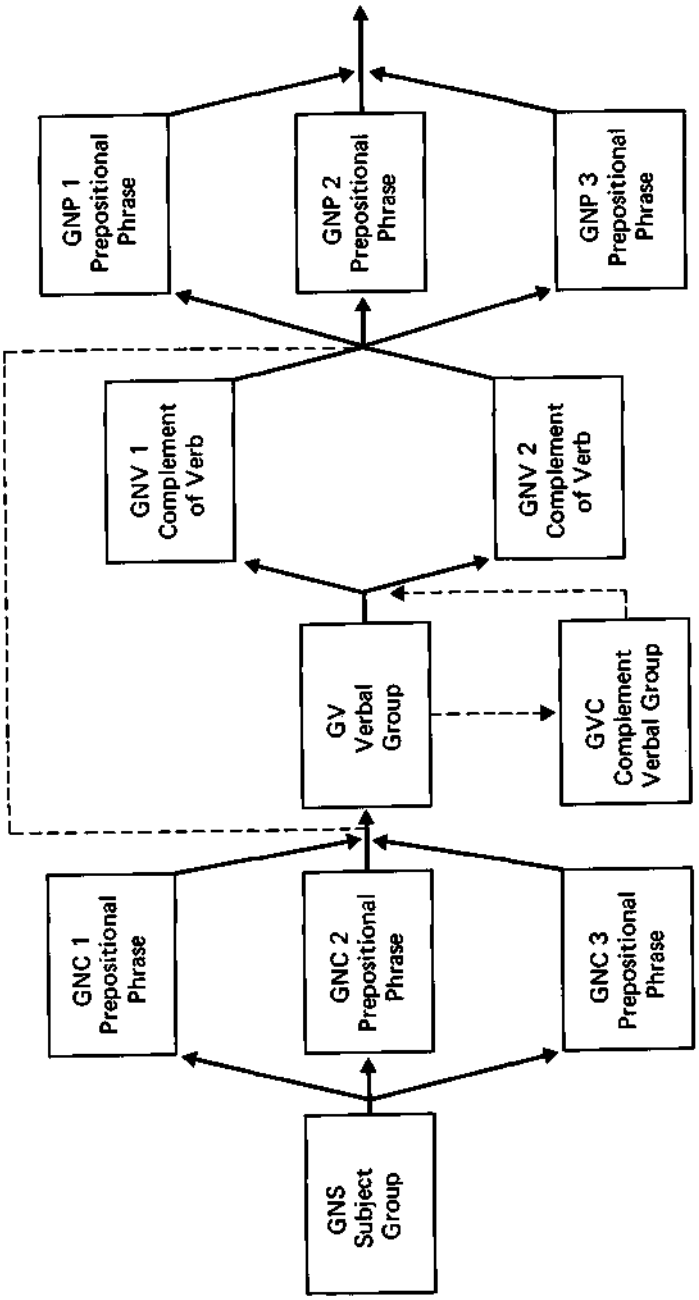

Figure 3. Basic model of a sentence as established for TITUS 4