FABS (FORMULATED ABSTRACTING): AN EXPERIMENT IN REGULARIZED CONTENT DESCRIPTION (FABS: UNE ETUDE POUR LA FORMALISATION DU RESUME SIGNALETIQUE)

Brian Harris
Linguistics Documentation Centre
&
Thomas R. Hofmann
Dept. of Linguistics & Modern Languages
University of Ottawa
Ottawa, Ontario K1N 6N5

ABSTRACT

A bilingual experiment is being conducted at the Linguistics Documentation Centre, University of Ottawa, into the elaboration of well structured formulary routines for making the writing of abstracts easier, at the same time standardizing and generally augmenting the information given in them. (A l'Informathèque de Linguistique de l'Université d'Ottawa, une expérience bilingue tente présentement, à travers un formulaire rigoureusement structuré, d'établir des procédés routiniers qui faciliteraient la rédaction des résumés signalétiques, tout en favorisant la standardisation et le nombre d'informations.)

CAVEAT

This is a preliminary report of research that has begun only recently (Nov. 1972). It is being conducted on a very small scale. Even so, not all of the work can be described within the space limit of this paper.

INTRODUCTION & PRINCIPLES

Harris (1971) has shown how traditional bibliographic descriptions -- the information and layout usually employed in bibliographies or on author-title catalogue cards -- are so regular that most of their syntax can be described by a formal context-free grammar. Some advantages of this regularity are:

(i) Interchangeability of catalogue cards and easy reading of other people's catalogues and bibliographies.

(ii) Clearly implied directions to the cataloguer/bibliographer as to what information to extract from documents; which vastly helps the maintenance of standards in the matter.

(iii) Fields that are well delimited for information retrieval systems.

In sharp contrast to this epitome of regularity stand content descriptions that are written in the form of running text: the typical example is abstracts. (By this definition we exclude from consideration uniterm, multiterm and facetted classification languages.) From the linguistic point of view abstracting remains freely creative; but consequently it lacks the advantages just mentioned.

Advantages (i) and (iii) above accrue to the user. So does (ii) eventually, but it first affects the documentalist. Those implied directions help the latter by telling him:

(a) What he must not fail to account for, even if it might not all appear of interest to him personally.

(b) The order in which to set it down, relieving him of many decisions about the expressions and punctuation to use.

Conversely, misunderstandings ommissions and garbling are reduced.

If the constraints of a well tried format make work easier, the corollary is that the operation should be faster and cheaper. Even if the format is not rigidly adhered to in practice, a training in it should habituate an intellect to seeking out requisite information and should leave suitable terminology ready to spring to mind to express it. The help that a tyro abstracter receives at present is usually vague general advice such as, "Pick out what is new" or "What is the thesis of the author?"

Working with and within a regulated framework does not render the task entirely mechanical. Nothing of value can be had without allowing the bibliographer to make judgements, and the primary judgement to be made is that of selecting salient information. Of course much research is going on into automating the selection process; however, the potential connection with automatic abstracting lies beyond the scope of the present experiment.

There is a spectrum of abstracts that ranges from the very brief 'indicative' abstract -- of which the abstract at the head of this paper is an example -- through the 'informative abstract' which "presents the conceptual content" of the document, to the critical review which may sometimes rise to the status of an important work in its own right.

FABS

Amongst the criteria laid down for the formulary to be used in this experiment were the following:

(i) It must be as general as possible in the sense that it can be applied to as many documents and areas as possible.

(ii) It must avoid areas of judgement that are open to wide disagreement. Value judgements are to be excluded, although they may be put into...

(iii) Additional comments which are allowed as a supplement to the formulated information in case the specifics of the formulary neglect some important aspect. Indeed these additional comments should provide the basis for expansion of the formulary itself.

(iv) For the same reason, the analyst should be allowed to use a term of his own within the formulary when he feels that none of those provided fits the document.

These criteria and a number of other considerations are dealt with at greater length in Hofmann (1972).

METHODOLOGY

As a first step, an initial formulary was drawn up in English. This primitive version is reproduced as Fig. 1. When it is compared with guidelines issued by a well-established abstracting agency (Fig. 2), it is seen to cover much the same ground, and seems to prescribe a fullness somewhere between that of the informative and the indicative abstract. The difference lies foremost in the help given to the abstracter, because at the same time as he is directed to seek information he is provided with ready-made terminology and syntax. We will turn later to consideration of the product.

By application of the guidelines quoted above, the formulary has been considerably expanded, both in its terminology and in its constituent structure. A recent English version is Fig. 3. The expansions come about when one or other of the two analysts working on the documents (L. Légaré and M. Gelbert) exceeds the formulary. They then discuss with a coordinator whether their addition is really necessary or whether they have failed to make full use of what the formulary already provides. If their addition is accepted as necessary, it is incorporated in an updated version. After five months of this trial and error, the formulary seems to have a chance of reaching satisfactory stability: that we would define as no further changes in the constituent structure and not more than one addition to the vocabulary per 1,000 documents analyzed.

Légaré is working on French documents, and while so doing is compiling a French translation of the formulary which will be published later.

As a rough and ready way of early evaluation, we have had the analysts do abstracts of documents that had already been abstracted in reputable journals, namely "Language and Automation" and "Language and Language Behavior Abstracts". Of course our analysts did not see the other abstracts before doing theirs. We then compared the items of information given in the paired abstracts. An example of this sort of comparison is Fig. 4.

Approximately 50 documents have so far been analyzed.

EXPANSION OF THE FORMULARY

The increased flexibility, subtlety and precision provided by the larger vocabulary in Fig. 3 is obvious. However, it also contains some deadwood in the shape of expressions that were in the original formulary (Fig. 1) but have not proved their worth. Sooner or later, when we have enough data to do it safely, we shall have to prune.

At this point an explanation of the notation used in Figs. 1 and 3 is called for.

{....} groups a set of terms out of which the abstractor may (in some cases must) chose one term each time be works through the formulary.

(....) surrounds sets that are optional. The abstractor should always consider whether they are applicable, but often they are not. Conversely, sets that are not in parentheses are obligatory.

[....] surrounds an explanation which is intended to guide the abstractor but is not for use in his text.

CAPITALS distinguish the literals, i.e. the terms themselves that are to be used.

Underlined lower-case letters are used for variables.

Variables: x, y, z for noun phrases, s for sentences, l for language names or types; / separates alternative variables.

The expansion -- a significantly increased complexity -- in the constituent structure reflects the complexity of the documents themselves and was demanded if one wanted to retain more than the very briefest of indicative-type information.

The sophistications are of three kinds:

(i) The split of the original sections II and III into II, III and IV so as to introduce certain information that is now provided for by III and take some of the functional load off II and IV.

(ii) The addition of several optional constituents: see the parentheses notation above. This gives more flexibility.

(iii) Footnoting, to accommodate the additional comments prescribed by the criteria. Footnotes can also be used for bibliographical references.

(iv) Most important perhaps, the introduction of the iterative mechanism and its notation, (see below).

ITERATION

A single straight pass through the formulary would still only permit a brief indicative abstract. Indeed one way to force brevity is to insist that it be used that way (cf. Fig. 5, no. 3).

For analysis in greater depth, however, it soon became clear that a way was needed to put in more information, and this without having to make the distinction between essential and marginal information that footnotes imply. On the other hand, the general aim of the project required that the syntax be kept formally simple and, as linguists would say, 'transparent' (i.e. marked overtly). The solution was to allow any amount of 'backtracking'; at the extreme one can do a complete 'da capo' from section IV to section I. So that instead of being restricted to the order

(i) I → II → III → IV

one can go

(ii) I → II → III
↳ II → III → IV

or

(iii) I → II → III → IV
↳ I → II
↳ II
↳ II → III → IV

and so on.

Each backtrack is marked by the special conjunction '&', or by a string of &s in which there is an additional & for each section backwards that one moves. Translated into this notation, the above three examples are rendered:

(i) I II III IV

(ii) I II III &&II III IV

(iii) I II III IV &&&&I II &II &II III IV

The form 'and' is reserved for normal intra-sentential use within a section.

Though this may sound formidable, we feel the product can be read without strain: see the examples in Fig. 5, where some help is given to the eye by paragraphing.

INTERIM CONCLUSION

So far as we are aware, no other research is being done in this direction. Our immediate aim is to arrive at a satisfactory formulary; that is to say one which is comfortable and helpful for the abstractor, readable and useful for researchers, easily parsed by computer. We are feeling our way within these constraints. Reduction in the cost and increase in the speed of abstracting are very important problem areas, but cannot be tackled until an ulterior stage of research and would require more substantial funding.

REFERENCES

HARRIS, B., (1971) "A Justification and a Suggestion for a Linguistics Documentation Language." Cahiers linguistiques d'Ottawa, no. 1 (Sept. 1971), p. 7-25.

HOFMANN, T.R. (1972), "A Content Bibliography". Unpublished.

FABS

Fig. 1: FIRST VERSION OF FABS FORMULARY

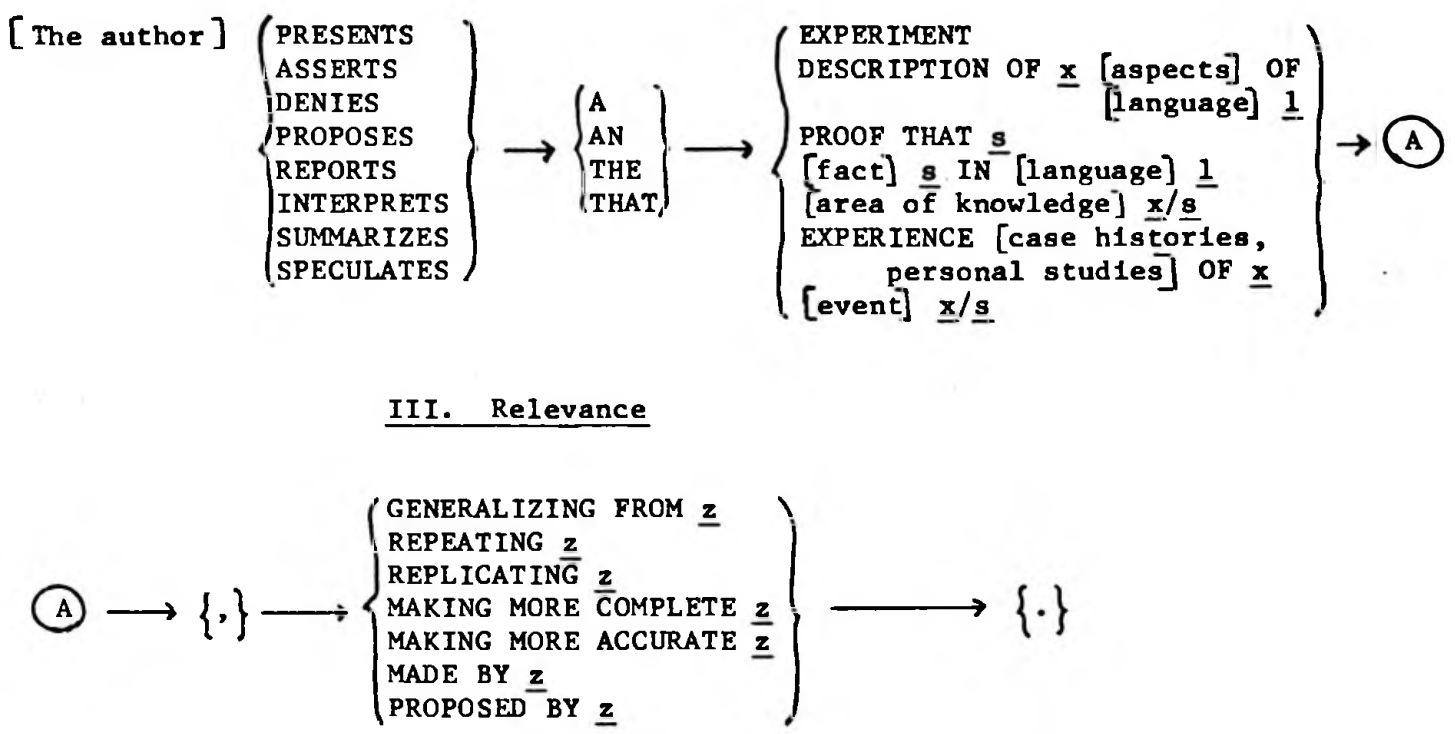


Fig. 2: TYPICAL GUIDELINES FOR ABSTRACTORS

Source: "LLBA"

INFORMATIVE ABSTRACT

An *informative abstract* presents the conceptual content of an article. Summarizing the essential ideas in an article, the abstract should answer the following questions:

1. What is the thesis of the author? What hypotheses or theories are presented?
2. How are the main hypotheses developed? What data are used? What methods of isolating data are used? Was novel methodology employed? Are the data qualitatively and/or quantitatively manipulated? What tests, scale, indexes, or other summarizing techniques are used?
3. What are the proofs or evidence relevant to the hypotheses?
4. What conclusions are drawn? Are the hypotheses, ideas, concepts, theories, etc., supported or rejected? What new relationships are found, old ones reaffirmed or rejected?

The *informative abstract* should show the meaningful, coherent relationship between the author's ideas and arguments; furthermore, it should enable the researcher to see the difference between one article and others on the same subject.

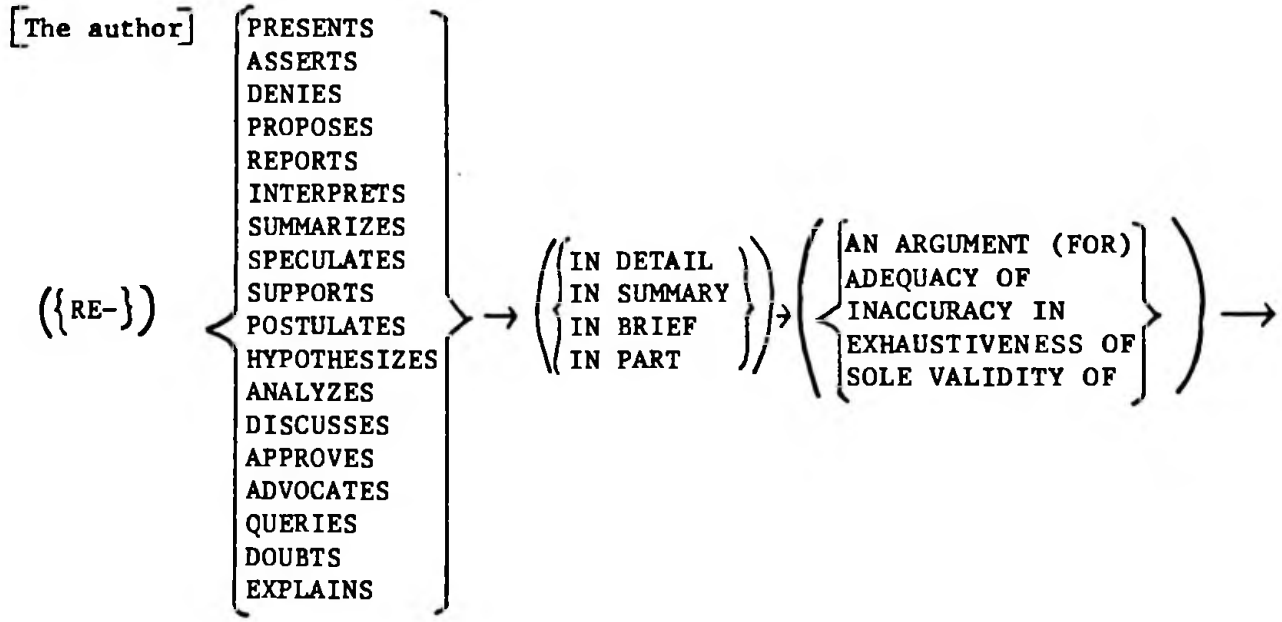
INDICATIVE ABSTRACT

Some articles (e.g., bibliographies, review articles, reports, and the like) cannot readily be summarized and require *indicative abstracts* which serve primarily as descriptive guides. An indicative abstract tells briefly what an article is about, what significant subjects it includes, and what its scope is.

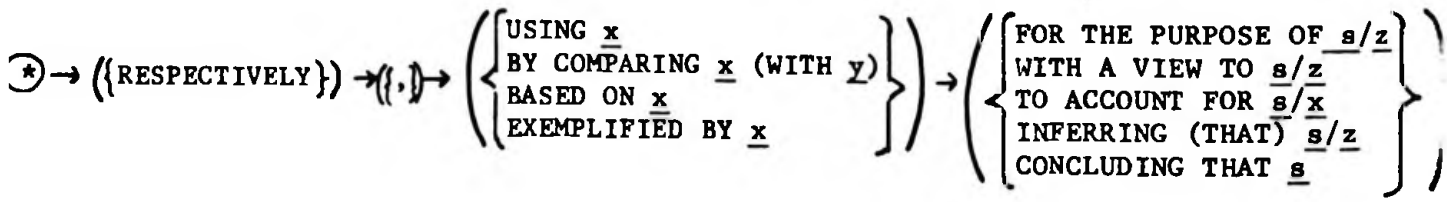
FABS

Fig. 3: FABS FORMULARY AT APRIL 1967

Section: I. Performative Verb

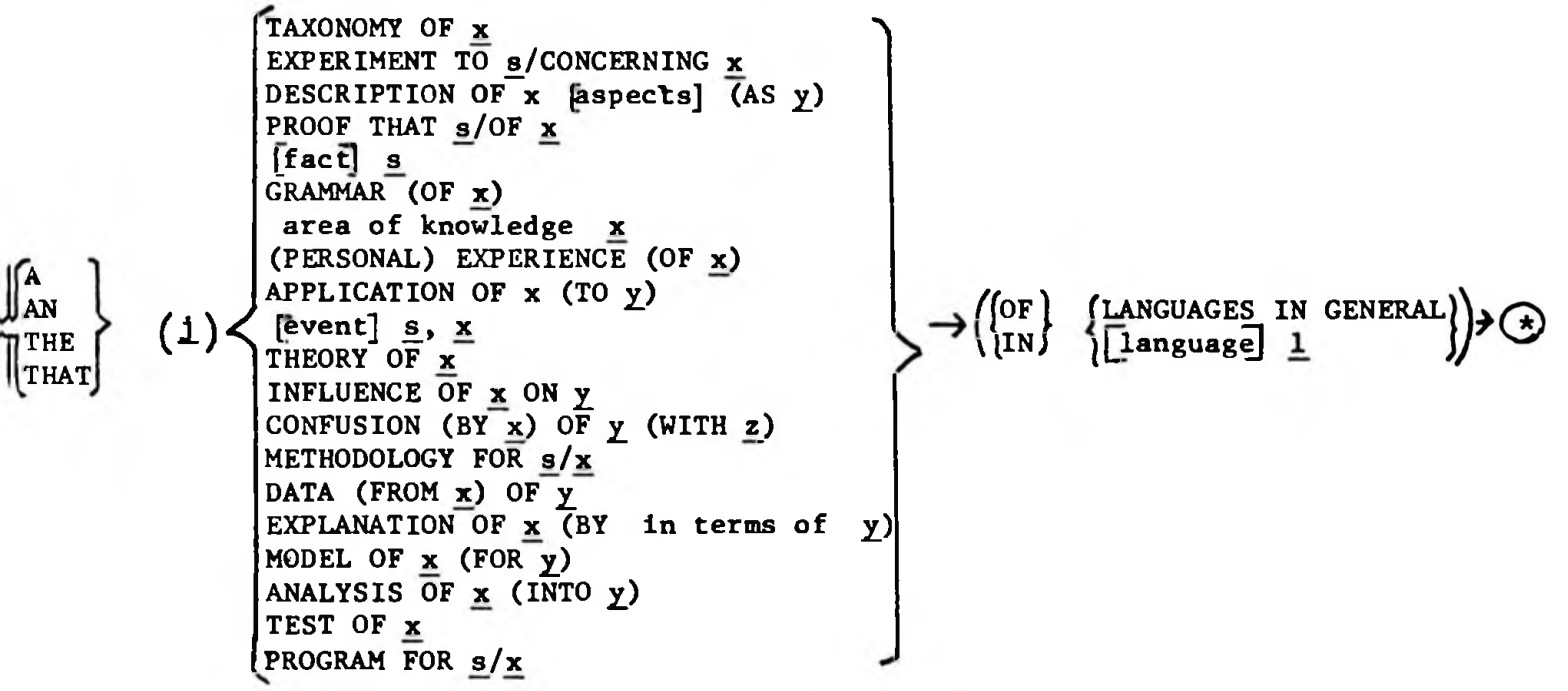


III. Means/Aims



FABS

II. Aspect/Type



IV. Relevance, Antecedents

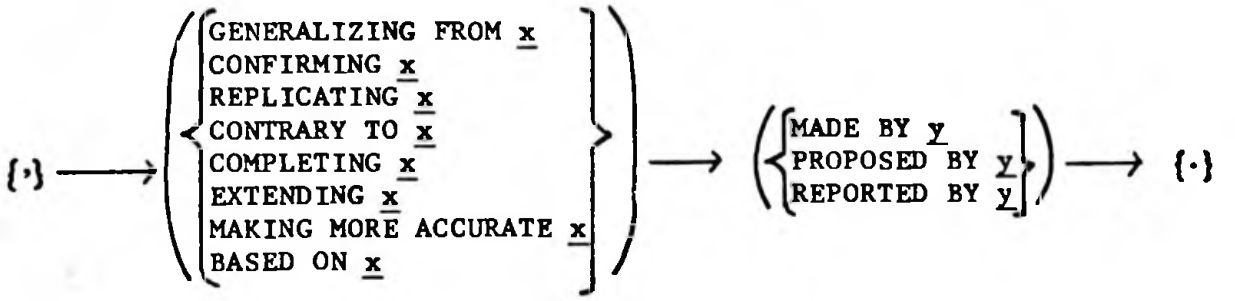


Fig. 4

Martin, E., "Truth and Translation", Philosophical Studies, v. 23 (1972), p. 125-130

FABS abstract

ASSERTS DESCRIPTION OF translation AS pairing of every sentence of one language with a sentence of another so as to maximize conditions of truth and falsity

&&&& DOUBTS classical deductive truth-theory EXPLANATION OF human semantic competence, CONTRARY TO Tarski

&&&& ADVOCATES rule-governed THEORY OF semantic competence GENERALIZING FROM generative syntax

&&&& ASSERTS THAT Tarskian THEORY OF truth is essentially the translation from object language to metalanguage it must contain.

(67 words)

Note: Expressions underlined have been lifted directly from the original text. 41 words are the same in both abstracts.

"Language and Language Behavior Abstracts"

[Our paragraphing and underlining - BH/TRH]

A discussion of the value of translation in framing a theory of meaning.

Production of a theory of truth appears to be at least a minimal demand in explaining human semantic competence although an extension of Tarski's classical truth theory would provide such a theory, it is not necessary that a standard deductive theory be used.

A theory can explain human semantic competence if it appeals to rules which produce desired instances of (T) and can be mastered by human beings.

Translation works by pairing every sentence of one language with a sentence of another so as to maximize conditions of truth and falsity, and gives us an understanding of the semantic roles of the sentence part of each language.

Translations between two languages essentially provide us with translations between an object language and the metalanguage it must contain.

(139 words)

FABS

Fig. 5: EXAMPLES OF FABS ABSTRACTS

1. Paillet, J.P. & Hofmann, T.R., "Assumptions of Integrative Semantics", in Integrative Semantics I (SRG Monographs), 1972.

   PRESENTS IN BRIEF an ANALYSIS OF semantics INTO contents of messages and contents of lexical items and praxis phenomena

   &&&& PRESENTS a formal DEFINITION of a C-net AS an oriented graph of relationships between labelled nodes (semantic atoms) and unlabelled nodes standing for individuals to be eventually related to a referent in a universe, FOR THE PURPOSE OF representing the content structures of discourse

   &&&& POSTULATES the ANALYSIS OF lexical items INTO subnets of C-nets

   &&&& PROPOSES a DEFINITION OF understanding AS successful integration of the current segment of a discourse into the C-net already built up.

2. De Possel, R., "Les Résultats obtenus depuis fin 1968 en reconnaissance des formes et en particulier en lecture automatique par le R.A.M.I.", T.A. Informations, 1972, no. 2, p. 22-24.

   REPORTS IN BRIEF[1] RESULTS FROM an automatic reader OF roman characters USING a contextual (digraph and trigraph) and semi-sequential model FOR PURPOSE OF reading print and computer-output microfilm.

   [1] Full text to appear in Automatisme.

3. Moinfar, D., "Défini et non-défini en persan", T.A. Informations, 1972, no. 2, p. 20-21.

   PRESENTS MODEL OF recognition of definite and indefinite functional values IN Persian USING a table which relates marker and content.

4. Vaissière, J., "Contribution à la synthèse par règles du français", T.A. Informations, 1972, no. 2, p. 1-16.

   PRESENTS IN SUMMARY[1] a GRAMMAR OF & a PROGRAM FOR generation of prosodic features IN French. BASED ON a phrase-structure syntax[2] & ON number of syllables per phrase, FOR PURPOSE OF speech synthesis.

   [1] of A.'s Thesis

   [2] that of CETA, Grenoble.