

A computational model for a text-to-speech translator in Italian

R. DELMONTE

1. The phonological processor

We shall describe the computer programme for the automatic translation of any text of Italian into naturally fluent speech, or Phonological Processor (FP), which is used to guide a speech synthesizer. The FP maps into prosodic structures the phonological rules of Italian. Structural information is provided by such hierarchical prosodic constituents as Syllable (S), Metrical Foot (MF), Phonological Word (PW), Intonational Group (IG). Onto these structures, phonological rules are applied such as the "letter-to-sound" rules, automatic word-stress rules, internal stress hierarchy rules indicating secondary stress, external sandhi rules, phonological focus assignment, logical focus assignment. The FP constitutes also a model to simulate the reading process aloud, and the psycholinguistics and cognitive aspects related will be discussed in the computational model: i.e. the FP simulates the operations carried out by an Italian speaker when reading aloud any text. To this end the speaker shall use the rules of his internal grammar to translate graphic signs into natural speech. These rules can be organized into two levels. At a phonological high level we have suprasegmental rules, and at a low level we have segmental rules. The latter are represented by three levels of structure, that is S, MF, PW and are governed by phonological rules which are meant to render the movements of the vocal tract and the coarticulatory effects which occur regularly at word level and at word boundaries.

The former are represented by one level of structures, the IG, and are governed by rules which account for long range phenomena like pitch contour formation, intonation centre assignment, pauses.

In brief the rules that the FP applies are the following :

- a) transcription from grapheme to "phoneme", including the most regular coarticulatory and allophonic phenomena of the Italian language;
- b) automatic word stress assignment, including all the most frequent exceptions to the rules as well as individuation of homographs, which are very common in Italian;
- c) internal word stress hierarchy, with secondary stress/es assignment, individuation of unstressed diphthongs, triphthongs, hiatuses;
- d) external sandhi rules, operating at word boundaries and resulting in stress retraction, destressing, stress hierarchy modification, elision by assimilation and other phenomena;
- e) destressing of functional words listed in a table lookup, which is accessed first since these are higher frequency words;
- f) pauses marked off by punctuation; pauses deriving from a count of PWs; pauses deriving from syntactic structural phenomena; comma intonation marking of parentheticals, appositives and similar structures;
- g) rules to restructure the IG when too long, more than 7 PWs, or too short, less than 3 PWs;
- h) Focus Assignment Rules (FAR), firstly to mark Phonological Focus, or intonation centre dependent on lexical and phonologically dependent phenomena;
- i) FAR to mark phenomena dependent on syntactically and semantically determined intonation centre.

From a general computational point of view, the FP operates bottom-up to apply low level rules, analysing each syllable of each word at a time, until the PW structure is reached; it operates top-down to build higher structures, the IG, where to apply high level rules by analysing each PW at a time and conjoining them into IG.

As far as phonematic transcription of Italian texts is concerned, there seems to be no such difficulties as for English. In fact letter-to-sound rules are relatively few and quite straight forward to describe. There are a number of exceptions and counter-exceptions to the rules which have to be specified, but

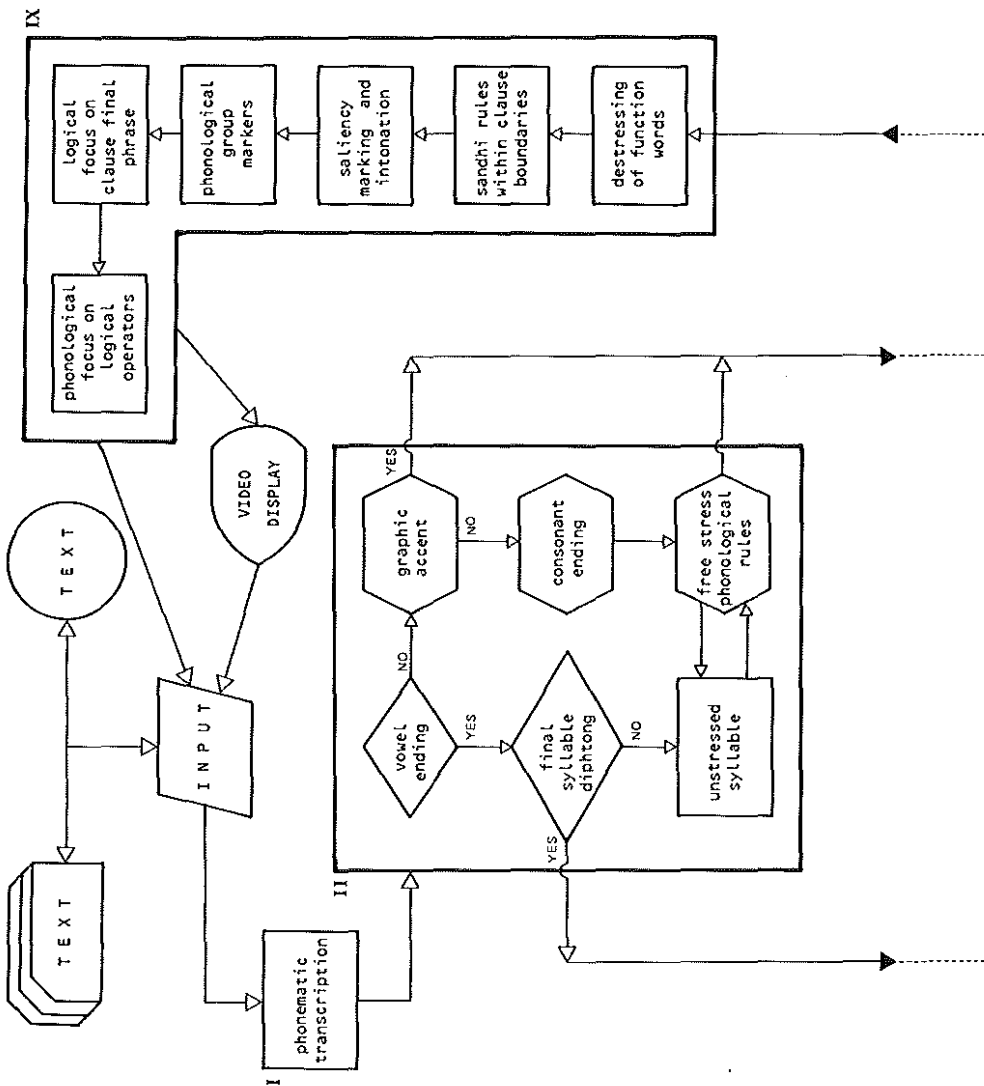


fig. 1
Flowchart of the prosodic translator

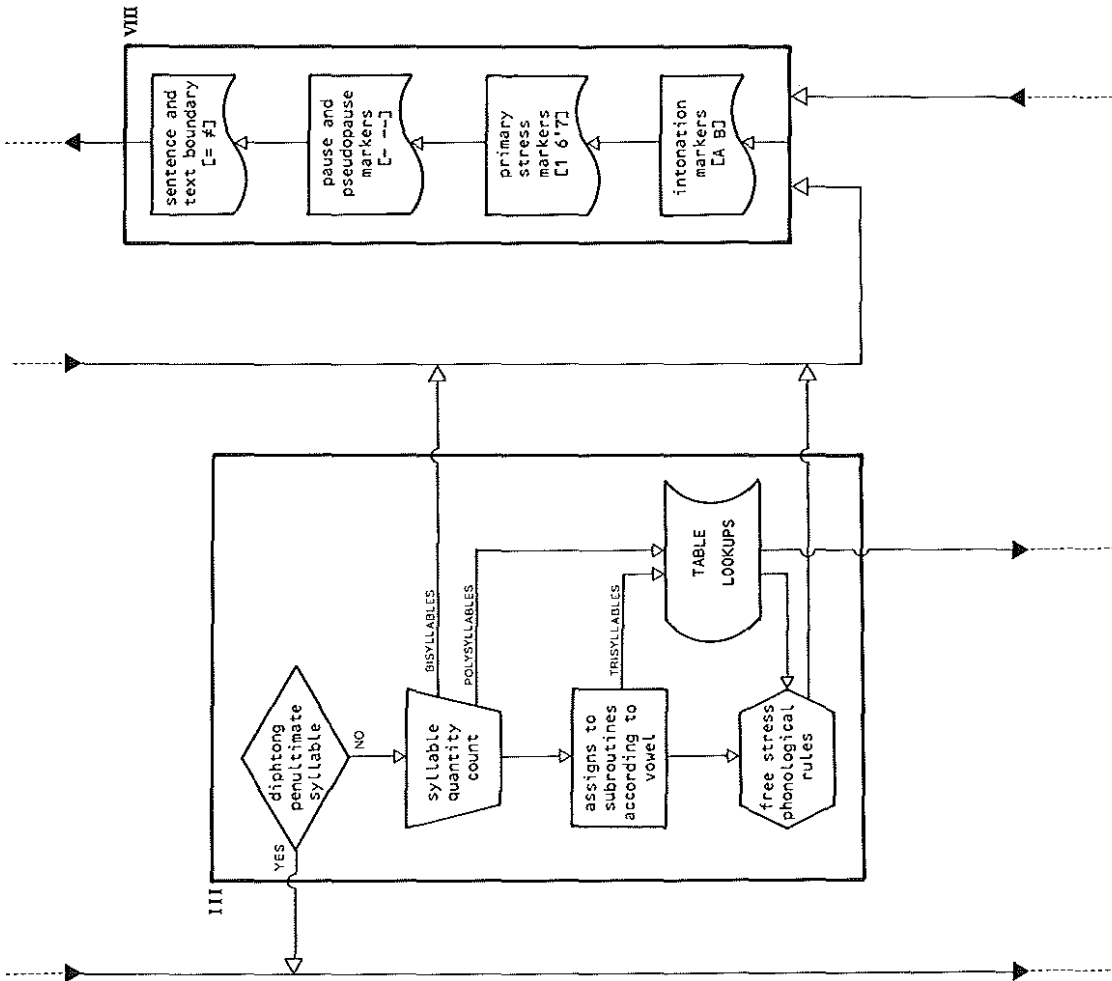


fig. 1 (continued)
Flowchart of the prosodic translator

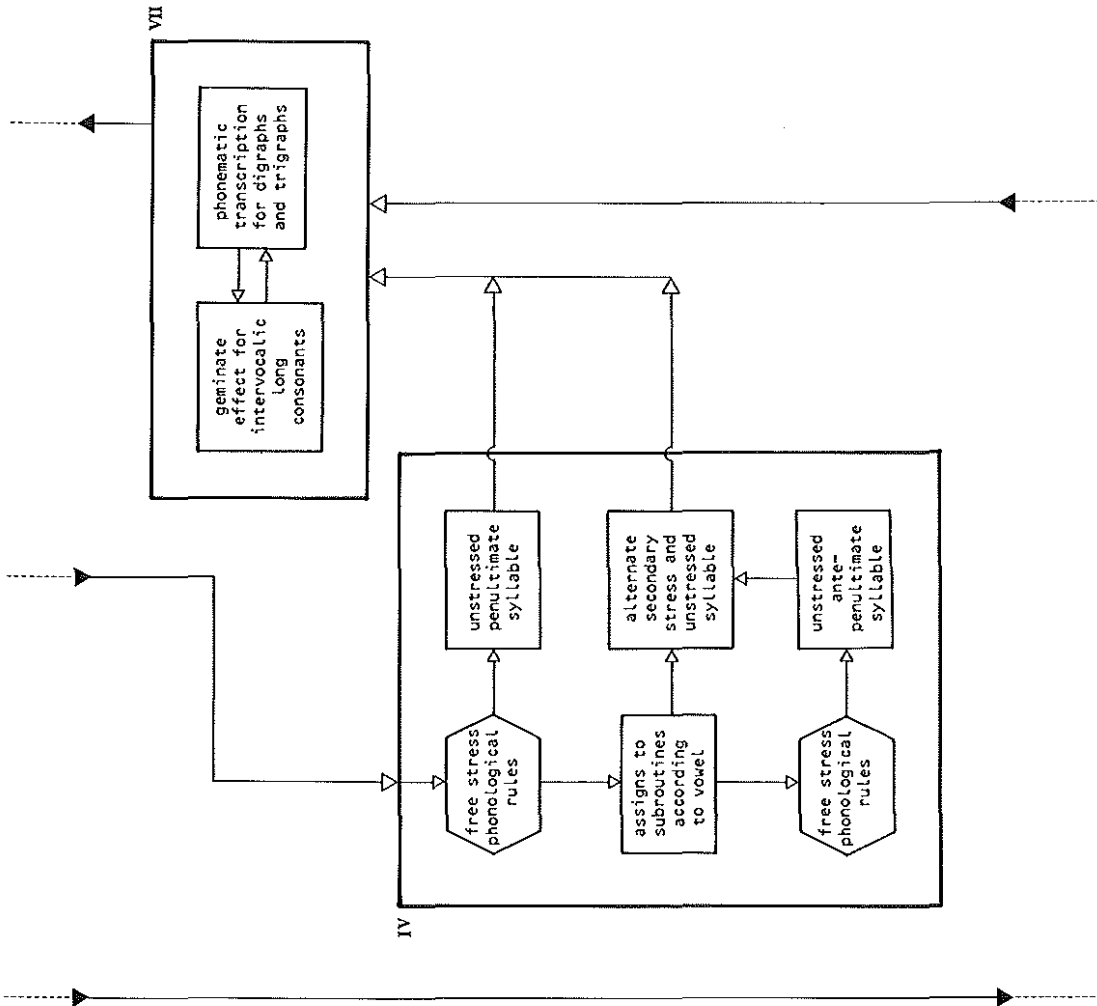


fig. 1 (continued)
Flowchart of the prosodic translator

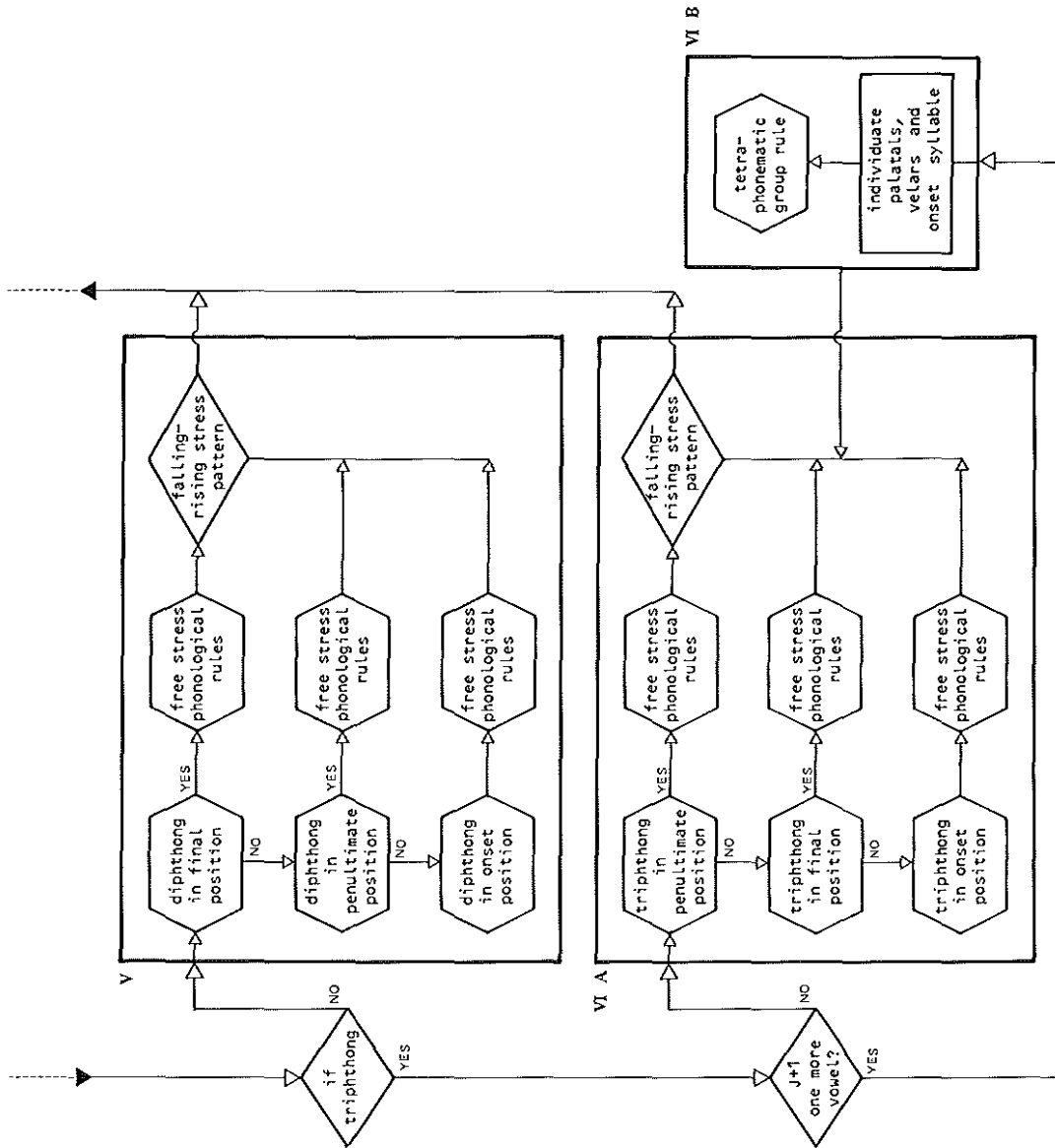


fig. 1 (continued)
Flowchart of the prosodic translator

no dictionary lookup seems to be needed. The main difficulties are created by digraphs and trigraphs which are ambiguous in that they can render both stops and palatals; some of the decisions concerning trigraphs must be delayed until stress has been assigned by word stress rules. Ambiguous graphemes include S and Z which can have both voiced/unvoiced phoneme pair. In the programme, the two graphemes have been treated together by the same set of rules operating conjunctively. Thus, a remarkable economy and simplicity has resulted; also, similar underlying phonological rules govern decision for voice assignment to S/Z. Letter-to-sound rules for [é,è] and [ó,ò] phoneme pair can be implemented. As to word-stress rules, it is our opinion that Italian speakers do not use directly morpho-syntactic information to assign word stress, but an ordered set of phonological rules to lexical items completely specified in a lexicon. Morphological information is restricted to a subclass of word types. Syntactic category information is required, also to tell homographs apart.

Speakers analyse fully specified lexical items by blocks of word-stress rules ordered sequentially, which address different types of words according to syllable structure. Words are made to enter each rule block disjunctively, that is each word either enters a block and receives stress, or is passed on to the next block. Exceptions are processed first: in our system they are part of the rules, so that each rule has its own exceptions also phonologically treated. In other systems they are regarded lexical exceptions and are simply listed in a table lookup and processed before entering the word-stress rule procedure (see for instance the system organized at CSELT by Prof. P.M. Bertinetto). No word can be sent back to steps of the algorithm already passed, that is there is no backtracking. The FP divides all words into two main classes according to lexical and morphological information: open class words, and functional or closed class words, the latter ones are destressed. Lexical open class words are made to enter into blocks of rules according to the following criteria:

- a) BLOCK I deals with words with graphic stress on the last syllable, like *carità*, *andrò*, *lunedì*, etc.; with truncated words, that is Italian words with consonant ending and foreign words; with monosyllabic words which can receive wordstress like *so*, a verb, or be treated as functional words like *lo*, article, clitic or pronoun but always destressed;

The following graphemes have been transcribed into symbols denoting "phonetic elements" :

k	=	CH	C+A,	O,	U	kk	=	CCH	CC+A,	O,	U	corresponding to phoneme	[k]
%	=	CI	CE	CI+Vowel		%%	=	CCI	CCE	CCI+Vowel		corresponding to phoneme	[tʃ]
j	=	GI	GE	GI+Vowel		JJ	=	GGI	GGE	GGI+Vowel		corresponding to phoneme	[dʒ]
/	=	SCI	SCE	SCI+Vowel								corresponding to phoneme	[ʃ]
=	GLI	GLE	GLI+Vowel									corresponding to phoneme	[ʌ]
=	GN+Vowel					XX	=	geminat	[z:]			corresponding to phoneme	[ʒ]
X	=	voiced S	[z]			&&	=	geminat	[dz:]				
&	=	voiced Z	[dz]										

And here are some exceptions :

GLICINE, ANGLIA, GERGLIFICO	where	GL = [gʲ]	not	[ʌ]
FARMACIA, LUCIA	where	CI = [tʃi]	not	[tʃ]
BUGIA, AEROFAGIA, NOSTALGIA	where	GI = [dʒi]	not	[dʒ]
SCIA	where	SCI = [ʃi]	not	[ʃ]

And here after is the Flowchart of the Rules for the transcription of graphemes S and Z which have both voiced/unvoiced phonemes.

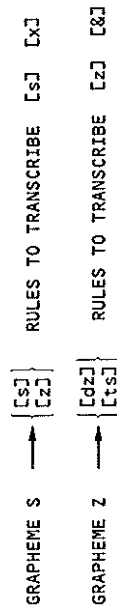
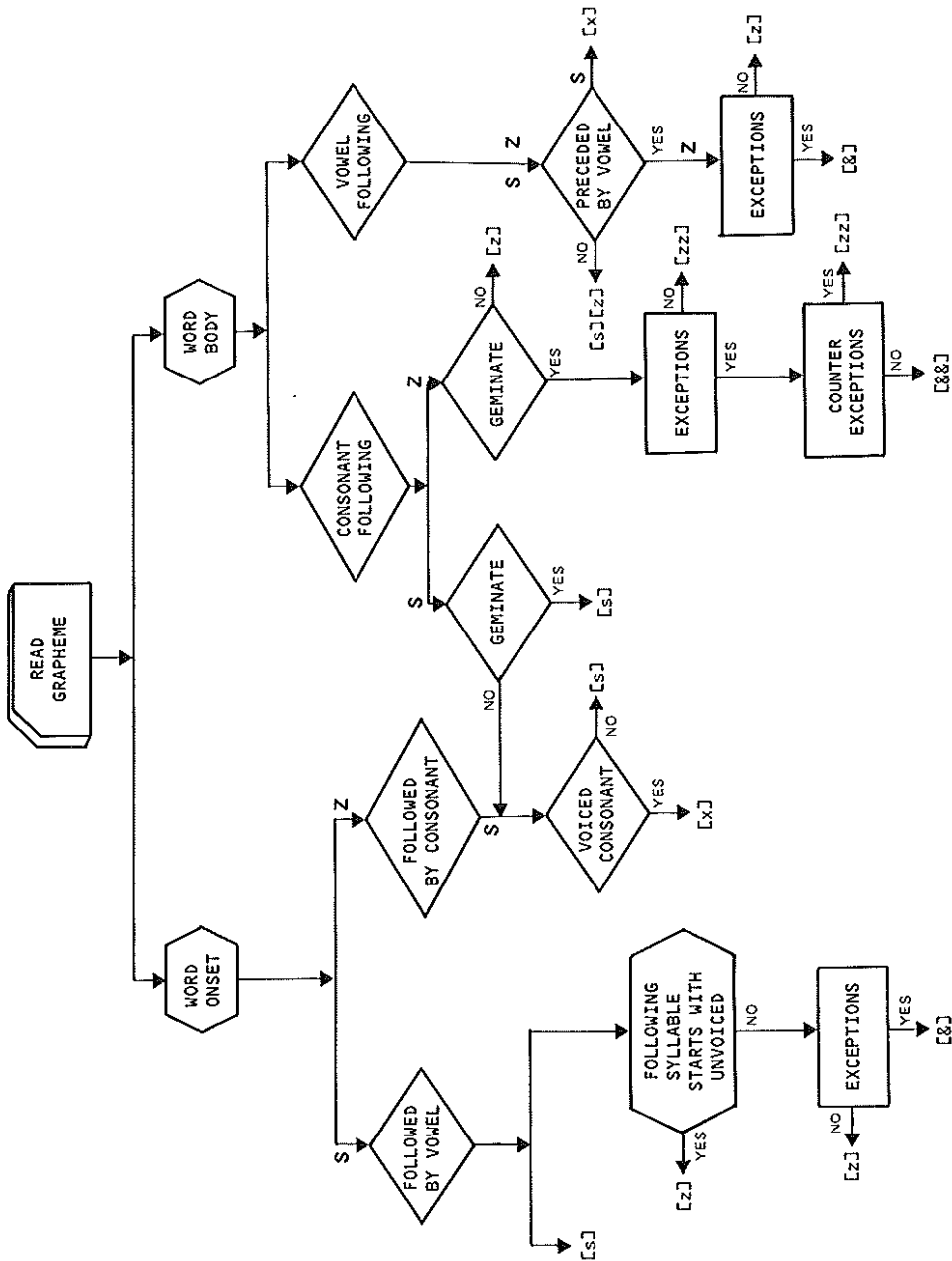


fig. 2
Letter-to-sound rules for s, z



- b) BLOCK II deals with bisyllabic words and applies to all words the first general word-stress rule which states that if a word has an heavy syllable in penultimate position it will receive stress on that syllable;
- c) BLOCK III deals with trisyllabic words and with all words ending with -ERvowel#, in which stress may result on the penultimate syllable if exception, and on the antepenult if regular;
- d) BLOCK IV deals with all words with more than three syllables, generally called polysyllables; it assigns morphological stress that is word-stress determined by morphological derivation processes based exclusively on suffixes. More work needs to be done on this section of Block IV. Information from the lexicon and the morphological analyzer is needed in order to divide adjectives of the type ending -ANO from verbs with the same ending: the two classes will receive stress on a different syllable position;
- e) BLOCK V with further subroutines deals with words either ending with a syllable containing more than one vowel, or with more than one vowel in penultimate syllable. Biphonematic, triphonematic or tetraphonematic vowel groups may result in diphthong, triphthongs, or hiatuses like *bugia*, *colloquiamo*, *acciaio*, *aiuole*.

Word stress rules take into account a series of phonotactic conditions, as well as the syntactic category of the word which is essential to the treatment of homographs and to verb treatment. In fact, Italian is very rich in homographs such as

ámbito, *ápriete*

where stress can shift from first to second syllable. Usually by varying the position of stress the meaning and the syntactic category will vary. Such words are included in a table lookup and receive stress from information derived from the morphological analyzer and the preanalyzer, which are used to take decisions about syntactic category assignment and about homography in general.

Another class of homographs, belonging this time to the same syntactic category is composed by such words as

ricòrdàti, *impicciàti*,

but also

príncipi, *àrbitri*, *balía*, *benèfici*, *malèfici*, *prèsidi*,
tùrbine, ecc.

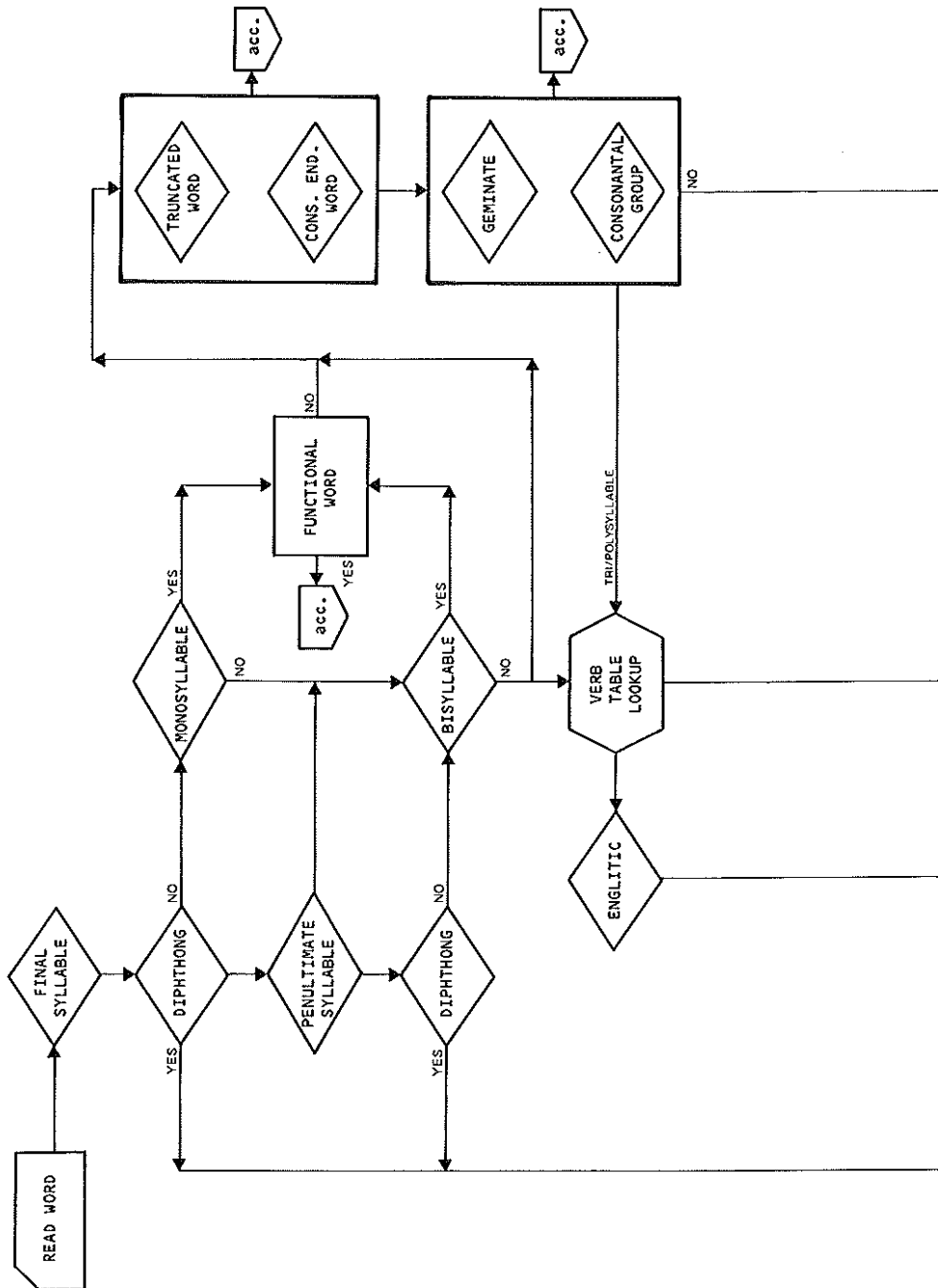
The first group is constituted by all verbs of the first conjugation and are processed according to the position they occupy in the utterance. If they come in first position or after a pause, it is assumed that they are cliticized imperatives and stress is assigned to the antepenultimate syllable. If they do not have that position in the utterance and an unstressed word or destressed word like an auxiliary precedes them they will be treated as past particles and stress is assigned to the penultimate syllable.

The second group of words must receive information from the syntactic pre-analyzer as well as from the morphological analyzer for instance where agreement decides to what category the word will belong as in *turbine*: this word will receive stress on the first syllable if singular, and on the second syllable if plural, meaning will obviously change, accordingly. Local contextual information is sufficient to take the most suitable decision. Where no such information is relevant, as in *balia* the lexicon and the subcategorization frame of the related verb in the sentence will decide where to assign stress. At present, the lexicon has been organized only in order to account for homography and syntactic category assignment.

Next come rules that take decision about internal words-stress hierarchy: they regard mainly secondary stress assignment but also an adequate definition of all unstressed syllables preceding and following the stressed ones, to suit coarticulatory and prosodic phenomena. To assign secondary stress the FP builds up the MP structure. This is done by counting the number of syllables preceding the stressed one. The rule states that the FP has to alternate one unstressed syllable before each primary or secondary stressed one. Restructuring may result in words with three or more than three syllables before the primary stressed one, as in :

fèlici'tà, autènci'tà, artificiali'tà, fotògra'fare,
cinemàto'grafico, rappresèntativa'mente, matematica'mente,
utilitarìstica'mente, precipitevolìssimevol'mente.

According to the number of syllables, two unstressed syllables may precede or follow the secondary stressed one. The Restructuring Rule for the MF takes into account performance factors which require that the number of secondary stressed syllables cannot be more than two when speaking at normal rate, but also that no more than three unstressed syllables may alternate stressed ones. To produce particular emphasis, i.e. when the word constitutes an utterance in itself, there may be



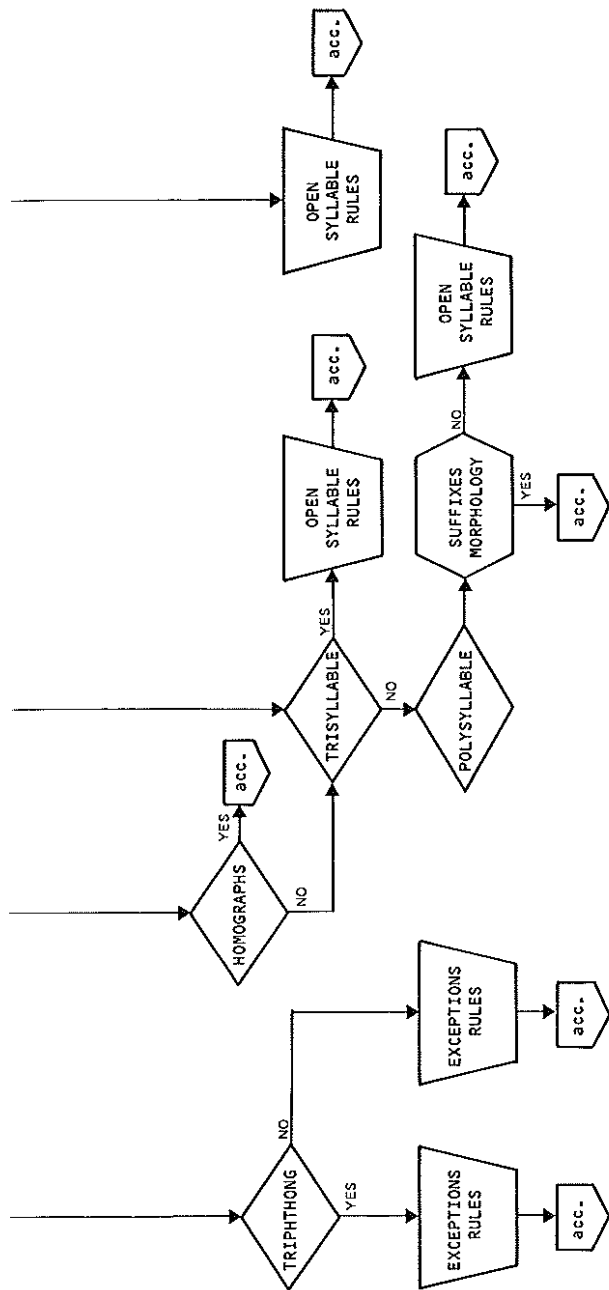


fig. 3
Flowchart of word-stress assignment rules

a remarkable increase in the number of secondary stresses to three or even to four secondary stresses in the same word, as in *precipitèvolissimèvol'mente*. As a result the speaking rate will slow down noticeably at values – number of syllables per second – well under the norm, to suit the speaker's attitude in a particular communicative situation.

Up to this point low level structures have been built by stress-sing content words and destressing functional words, which have become proclitics and are joined to the first stressed word on their right to build PWs, as in *dalla nostra parte* (on our side). High level rules localize punctuation pauses and start to apply external sandhi rules, which may elide a vowel, as in *la famiglia Agnelli, il maré è molto agitato*; or produce a schwa-like vowel as in *hannø interesse, è incredibile*; retract primary stress as in *'dottor 'Romolo, 'ingegner 'Rossi*. In the latter case, stress rules have to move back primary stress and to unstress the remaining syllables.

It is essential to apply these rules in this phase, because intonation centre may only be assigned to primary stressed syllables. Exceptions are represented either by auxiliaries which can assume the role of lexical verbs as in *oggi non ci sono* (today I'm not there), *ho chiesto ma non ce l'hanno* (I asked but they haven't got it); or by clitics and adjectives which can become pronouns as in *non ci vengo con te* (I don't come with you), *preferisco quella* (I prefer that one).

External sandhi rules are highly sensitive to syntactic closure phenomena, which can be adequately treated only if information is provided by a syntactic parser and a semantic analyzer. These rules specify as a general condition for their application the non existence of a pause between the two adjacent words to which they will apply. In particular, the two words can belong to two major constituents or to the same major constituent, but neither of them can be dominated by an F node different from the one that dominates the second word, nor the second word can be attached to the higher F node. In the following examples the theory of syntactic closure should account for the fact that the final constituent cannot possibly be attached to the major constituent on its left: in this way a pause could be assigned before it, and external sandhi rules will be prevented from applying :

- a) *Oggi mi ha presentato ai suoi amici/Ignazio.*
- b) *Mi hanno chiesto di non andare/espressamente.*
- c) *Ti avevo detto di portarli tutti/insomma.*

Syntactic closure phenomena, which can be dealt with by a sophisticated parser, are tightly related to IG construal rules. These rules transform one or more PWs by joining them into an IG, which shall receive the intonation centre by Focus Assignment Rules (FAR). IG construal rules should intuitively build well formed IGs. General well-formedness conditions could be established so that phonological facts reflecting performance limitations as well as syntactic and semantic structural facts can be adequately taken into account. These conditions are formulated as follows.

CONDITIONS A :

determined by intrinsic features of the functioning of memory and of the articulatory apparatus or vocal tract, which impose restrictions on the length of an IG – length is defined in terms of number of constituents, i.e. PWs, to be packed into an IG; this number can vary with the speaking rate and other performance factors which are strictly related to temporal and spatial limitations of the language faculty;

CONDITIONS B :

determined by the need to transmit into an IG chunk of conceptual information concluded in itself and related to the rules of the internal grammar.

Construction Rules referring to Conditions A will first base their application on punctuation, assigning main pauses for each comma, full-stop, colon, semi-colon detected in the text. Restructuring may then take place according to the number of constituents present in each IG; if less than three, the IG is too short to stand on its own and it will be joined to the preceding one. If more than seven PWs, and the utterance is not yet ended, two IGs will result according to phrase structure as analysed by the grammar component. Provisionally, contextual information based on syntactic category is used: functional words are destressed and are treated as proclitics to be joined to the first following PW. To satisfy Condition B, the grammar component has to provide syntactic and semantic information to the FP.

The two already implemented grammatical subcomponents of the system, though in need of improvements and additions, are the Lexicon and the Morphological Analyzer. The former is made up of the following entries :

1. Fully specified entries
 - 1.1. Functional words;
 - 1.2. Base verbs;

1.3. Base nouns and adjectives;

1.4. Past Participles;

2. Stems and affixes

It contains internal lexical procedures to build up new words of the verb, adjective and noun category. The affixes, together with inflections for verb analysis, are organized in a morphological analyzer, which is the processor analysing the input word. Affixes, i.e. suffixes and prefixes are joined in graph structures, so that once the processor enters a graph tree it can only continue along one of its branches. When the root of the word is reached the processor also has reached a vertex and can continue its search within fully specified entries. Morphological information is thus made available accordingly: it can include number, gender, tense and syntactic category. The search can continue afterward if the word analysed contains in its onset one or more prefixes. Stems, about 200, are very useful to speed up the search analysis: its major drawback is that it loses in syntactic and semantic information content, which will have to be recovered elsewhere.

Syntactic category/ies information is listed beside each fully specified entry: they/it are all outputted with the word being analysed if the search has been successful. Otherwise, only branching depth and morphological types of the affixes being detected is provided. This applies also to stems.

The final choice among the various syntactic category labels provided by lexical and morphological analysis will be decided by a pre-analyzer which is being implemented at present. It is organized in a transition table from one syntactic category to the other of the 21 listed in the lexicon and the morphology. Cross-comparisons are being computed by the pre-analyzer and syntactic contextual restrictions are made to apply locally so that a first approximate result can be obtained as regards homographs belonging to different syntactic categories. The remaining ambiguities will be solved by the syntactic parser to be implemented.

Two different kinds of FAR are distinguished by the FP: marked and unmarked ones. The former are dependent on syntactic and semantic structural information; the latter are dependent on phonological and lexical information: thus we have Logical Focus and Phonological Focus. Phonological information is used to account for utterances such as simple declaratives, imperatives, wh-questions, yes/no questions, echo questions where

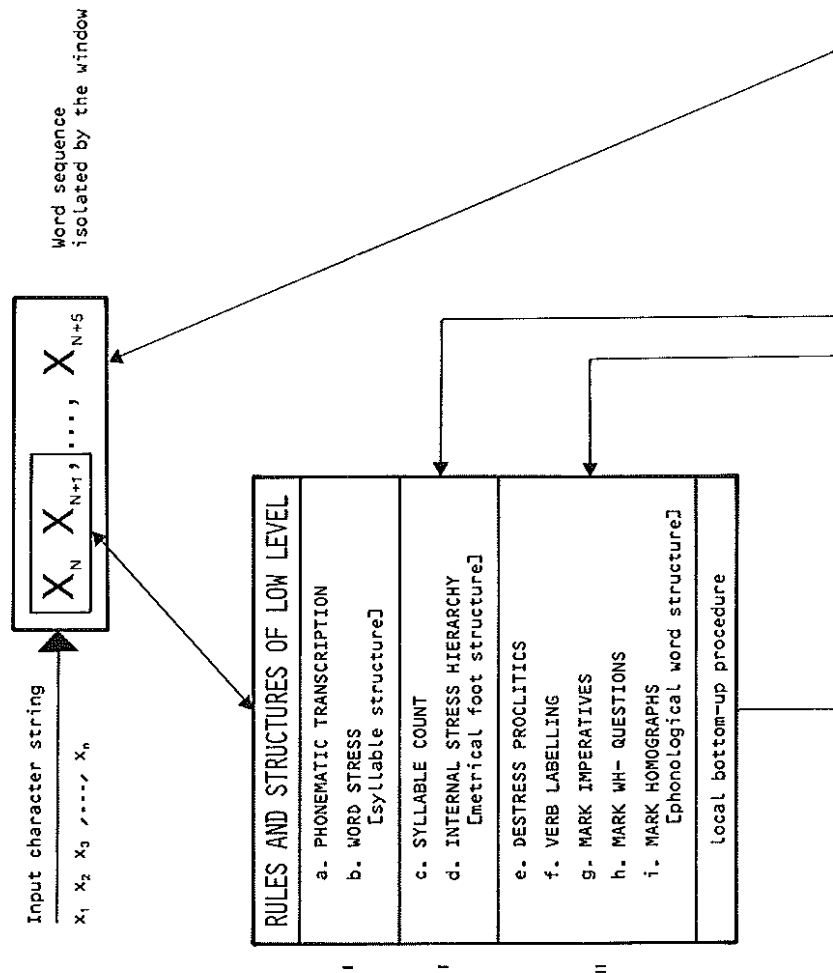
IGs can be built without structural information. Exceptions are wh- questions where extraposition of PP from NP are in act, or questions with words like *perché*, where focus is assigned to the PP or to the question word; also in imperatives, when a transitive verb is used, focus goes to the Object NP. In the normal cases, the Nuclear Stress Rule can be made to apply in a straightforward way, as follows: "Within an IG reduce to secondary stresses all primary stresses except the one farthest to the right".

Lexical information is required to label verbs and is passed on to the phonological component in order to assign focus to wh- questions and imperatives. It is also essential in order to spot logical operators which induce emphatic stress and attract the intonation centre of the utterance in their scope, shifting it to the left of the utterance, to the first PW following them. These lexical items are quantifiers and intensifiers such as MORE, MUCH, ALL, ALSO, ONLY, TOO, MANY, etc., which modify the semantic import of the utterance. As to the presence of Negation operators, in case the utterance is composed of more than one proposition, they can move the focus to the following one.

The second set of FAR, the marked ones, shall assign Logical Focus according to structural information relatively to those constituents which have been displaced or moved to the left/right of their predicates. For instance, left dislocation and topicalization will have to receive a different phonological treatment from normal declaratives, as follows :

- a) *La donna voleva fuggire*
- b) *La fronte voleva accarezzare*
- c) *La mamma voleva scrivergli*
- d) *Alla mamma voleva scrivere*

where the underlined constituents are assigned focus because in (a) the subcategorization frame of the verb *fuggire* will inform the parser that it is intransitive and takes no OBJECT; whereas *accarezzare*, a two-place predicate needs a direct object to produce a grammatical sentence, hence there is a topicalized NP; *scrivere* a three-place predicate takes a human SUBJECT and an inanimate OBJECT, hence *la mamma* is taken as subject in (c); it can also take an indirect object as a third argument which must be human, and this is fulfilled by *alla mamma*, the PP left dislocated in (d).



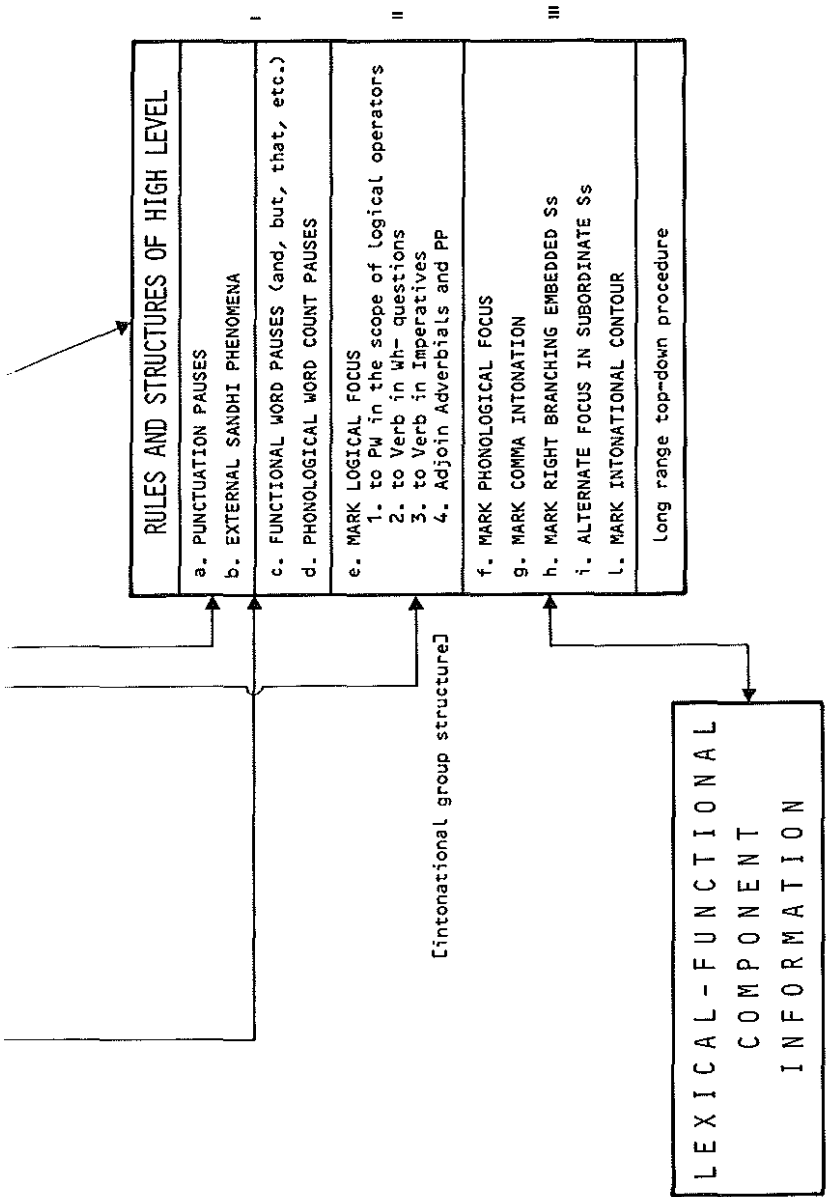


fig. 4
Computational linguistic model to simulate the process of reading

2. The Computational Mechanism

So far, we have described the rules of which the FP is equipped. We shall now deal with the psycholinguistic and cognitive aspects of the FP which, as we already said, is a model to simulate the process of reading aloud any text. From the previous description, it would seem that a speaker analyses the utterance proceeding at first bottom-up, until all low level rules have been applied to the structure of PW; subsequently, he should apply high level rules and he should build up IGs operating top-down.

In fact, the two procedures will have to interact at certain points of the utterance so that both low and high level rules will be applied contemporarily and fluent reading aloud will result. Whereas the speaker applies low level rules each time the graphic boundary of a word is reached, to apply high level rules he will have to wait for the end of an IG, which could be determined phonologically, lexically, syntactically and functionally or semantically. Intuitively, as he proceeds in the reading process, the speaker will stress open class words and destress closed class ones; he will assign internal stress hierarchy and at the same time he will look for the most adequate sites to assign main pauses; he will apply external sandhi rules, modifying, if required, the previous internal stress hierarchy; he will build up pitch contour according to the intonational typology appropriate to the utterance he is producing; intonation centre may result shifted to the left if he encounters logical operators, or to the end of the utterance, provided that it is not a complex proposition with embedded and subordinate structures in it.

To carry out such an interchange of rule application between the two levels of analysis of the utterance, the FP shall have to jump from one level to the other if need be. It will then be provided with a window which enables it to do a look-ahead in order to acquire two kinds of information: the one related to the presence of blanks or graphic boundaries between words, and the other related to the presence of punctuation marks. The window we have designed for the FP enables it to inspect five consecutive words, but not to know which of these words will become the head of a PW or a PW itself, at least not before low level rules will apply. The function of the window is then limited to the individuation of possible sites for punctuation pauses. But this is also what a reader will probably do while reading the text: as a matter of fact, he will surely want to

know how many graphic words are left before the end of the utterance is reached. Graphic information provided by the window is vital then, both for low level and high level rules application.

As far as low level rules are concerned, the local bottom-up procedure is well justified since the reader will want to know first if the word ends with a graphic stress mark, assigning word stress immediately; if this is not the case, he will turn to the penultimate syllable, which is the site where Italian words stress assignment is decided, and he will carry out syllable count if needed. Word stress rules will apply and internal stress hierarchy will be assigned.

The main decision to be taken before high level rules may start to apply regards pauses. As we said before, visual information may guide the reader together with phonological decisions previously taken. But quantitative count of words still left to process is only the first criterion, which shall have to be confirmed by qualitative analysis on a structural level. Structurally assigned pauses shall account for subordinate, coordinate propositions as well as embedded ones. Whereas comma intonation will have to be assigned to appositives, parentheticals and non-restrictive clauses, subordinate propositions may be assigned Focus. In this context syntactic closure phenomena are of paramount importance to the interrelation between the processor, syntax and semantics. Graphic information – the presence of one or two commas in the utterance – may thus receive two completely different interpretations: the FP shall have to individuate subordinate clauses which are usually preceded by adverbials, linkers or conjuncts such as *se*, *quando*, *sebbene*, *perchè*, *benchè*, etc. which cause temporary information storage and a suspension of RAF application. Focus goes to the subordinate only if it comes at the beginning of the utterance and is not a proposition of the kind of concessives, consecutives, conditionals, adversatives which are easily detected from the kind of conjunct introducing them.

As far as embedded clauses are concerned, waiting for the complete parser to be activated, the FP operates only through the individuation of verbs and of complementizers. In particular, the presence of *che* may induce a pause only if the embedded clause is right-branching. Completive, like infinitives and indirect questions, as well as restrictive clauses do not require a pause unless a lexical subject is present.

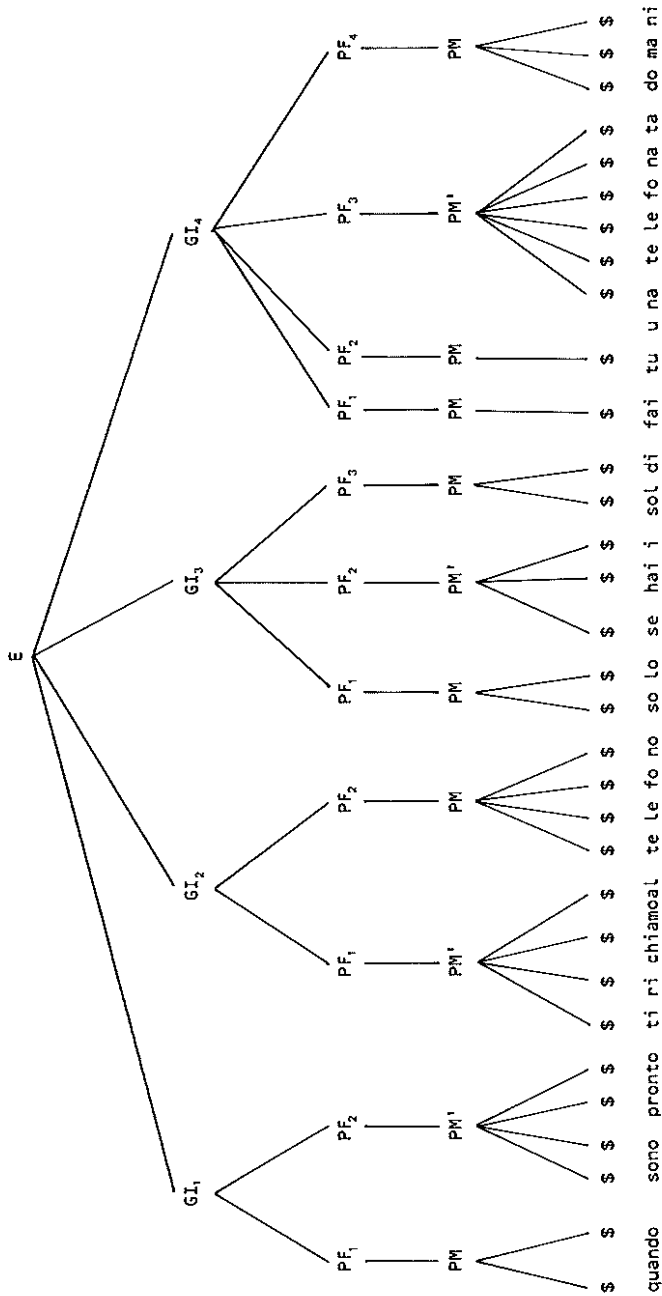


fig. 5
Phonological structure of utterance analysed by the FP

3. Higher level Phonological constituents and rules

We shall discuss hereafter only some of the theoretical details involved in the specification of our model. As already mentioned elsewhere, the phonological processor (FP) builds its own prosodic structures separately from syntactic concurrent processes. Information from the other grammar components is required only when the level of Intonational Group (IG) is reached, that is at the higher level. In this case phonological information is not sufficient and syntactic and semantic structures shall have to be computed by the other grammar components, and supplied to the FP. The theoretical proposal which in our opinion will suit best our performance oriented processor is the lexical functional one known as LFG, particularly because it allows direct syntactic and semantic coding on surface structure, without the intervention of an intermediate level of transformational rules. LFG computes semantic information from two basic subcomponents :

1. a lexicon, where each entry is completely specified so that only fully inflected and morphologically complete words are entered into phrase markers. Each entry has associated subcategorization features : lexical items subcategorize for grammatical functions such as SUBJECT, OBJECT and so on, and not for constituent structure categories. Each entry exerts selectional restrictions on a subset of its subcategorized functions; the predicate-argument structure of a lexical item lists the arguments for which there are selectional restrictions. These conditions are listed in a lexical form in which arguments are paired with functions, as well as grammatical function assignment which lists the syntactically subcategorized functions.
2. context-free rules to generate syntactic constituent structures. The combination of the descriptions will result in a constituent structure and a functional structure which represent formally the grammatical relations of the utterance analysed.
3. Functional relations intervening between predicate argument structure and adjuncts or complements, subcategorizable or not, are determined both by a theory of control which is an integral part of the LFG, as well as by a theory of syntactic closure which is essential to a theory of performance (see Delmonte, 1984).

Before going into the description of Focus Assignment Rules (FAR) we shall formulate the Rules of IG Construal, diffusedly exemplified elsewhere (Delmonte, 1983), but also commented further on :

RULES OF IG CONSTRUAL

1. Constituents moved by dislocations, clefting, extrapositions, and raising, obligatory form at least one IG;
2. Starting with the first PW of an utterance, join into one IG all PWs until you reach :
 - 2.1. the last constituent of a VP in Wh- questions and imperatives;
 - 2.2. the last constituent functionally controlled by a VP, i.e. an argument or a clause; complements and adjuncts functionally controlled by the main subcategorizable functions, according to syntactic closure attachments;
 - 2.3. the last constituent anaphorically controlled by a supraordinated clause where the matrix Subject appears, control is expressed at functional level by thematic restrictions.

We can now distinguish between two kinds of FAR, marked and unmarked ones. Unmarked FAR are dependent on phonological and lexical information and give rise to Phonological Focus; marked FAR are dependent on syntactic and semantic structure information and give rise to Logical Focus (see also Gueron, 1980).

Phonological information is used to account for utterances such as simple declaratives, imperatives, wh- questions, yes/no questions, echo questions, where the Nuclear Stress Rule (NSR) can be made to apply in quite a straightforward way. The NSR (see C & H), can be reformulated as follows: "Within an IG reduce to secondary stresses all primary stresses except the one farthest to the right", as in

(6) *Jáck stúdiés sècondáry éducatíon*

which is derived from an underlying representation where word stress is assigned phonologically by word stress rules,

(7) *Jáck stúdiés sècondáry éducatíon.*

The NSR for English works in the same way for Italian, as in:

(8) *Nella scúola súpériore, Giórghio non stúdia a sùfficiénza.*

Lexical information is required to label verbs and functional words to be destressed by the phonological component, and is used in order to assign focus to wh- questions and imperatives, as in :

- (9) *Che tipo di libri scrive la persona che hai salutato ieri?*
 (10) Smettila di far tutto quel baccano quando leggo un libro.

Lexical information is also essential in order to spot logical operators which induce emphatic intonation and attract the intonation centre of the utterance in their scope, usually shifting it to the left. These lexical items are words such as NO, MORE, MUCH, ALL, ALSO, ONLY, TOO, etc. (see Jackendoff, 1972). These words modify the semantic import of the utterance and attract the intonation centre to the first PW in their scope, or in the case of negation, which can modify the whole utterance, they move the focus to the following proposition, as in :

- (11) *Anche Giorgio racconterà una bella storia.*
 (12) *Gli studenti hanno fatto molti esami nella sessione estiva.*
 (13) *Il bandito non ha ucciso il poliziotto.*
 (14) *Il bandito non ha ucciso il poliziotto, ma la persona alle sue spalle.*

A second set of FAR, the marked ones, shall assign logical focus according to structural information provided by the parser which shall have to detect those constituents which have been displaced and moved to the left. Within LFG framework, this information is supplied by the augmentation which is worked on the context-free c-structure grammar, by means of functional descriptions which serves as an intermediary between c-structure and f-structure.

Long distance phenomena like questions, relatives, clefting, subject raising, extrapositions and so on are easily spotted by the use of variables which can represent both immediately dominated metavariables – specified as subcategorization features in the lexicon – and bounded domination metavariables, the nodes to which they will be attached are farther away in the c-structure, and are empty in f-structure representation. Focus is thus assigned to the OBJECT argument of the verb in :

- (15) John has some books to read.
 (16) I have plans for tonight.
 (17) It is the cream that I like.
 (18) Ann I love.

Other structures like relatives, tough movements, subject raising behave differently from English: in Italian focus may be assigned phonologically, mostly, as in :

- (19) *Ho visto il vento muovere le foglie.*
 (20) *E' facile per Bruno conquistare Maria.*
 (21) *Maria è facile per Bruno da conquistare.*
 (22) *Elena ha lasciato istruzioni che Giorgio eseguirà.*
 (23) *A Maria è piaciuta la proposta che le ha lasciato Gino.*

Focus marked with a discontinuous line is optimal and emphatic and lexically derived by the use of the verb "piacere", but it is still different from focus marking in the corresponding English utterance (see Stockwell, 1972).

We can thus formulate the following :

FOCUS ASSIGNMENT RULES

1. Questions

1.1. in wh- questions focus is assigned to the VP; adverbials and other adjuncts are joined to the Verb and receive focus;

1.1.1. according to the functional role assumed by the arguments of the verb, focus can be assigned to the NP argument acting as Agent SUBJECT;

1.1.2. if extraposition of PP from NP are in act, or a question word like "perché" is present, focus is assigned to the PP;

1.2. in yes/no questions and echo questions, assign Focus phonologically;

2. Imperatives

Focus is assigned to the Verb according to predicate argument structure; adjuncts are joined to the Verb and receive Focus;

3. Declaratives

3.1. if there are arguments displaced to the left of the SUBJECT, focus will be assigned to the last constituent farthest to the right by NSR; topicalizations, clefting and some kind of extrapositions attract focus to the displaced argument;

3.2. if there are propositional complements, focus will be assigned again by NSR;

- 3.3. parentheticals, appositives, non-restrictive relatives will be assigned comma intonation;
- 3.4. with multiple embedded structures, focus assignment is conditioned by the presence of a lexical SUBJECT non anaphorically controlled by the SUBJECT of a supraordinated proposition; if so, more IGs will be built and more than one focus will result;
- 3.5. adjuncts can be assigned focus only if attached to the last argument controlled by the predicate, according to syntactic closure phenomena.

4. The grammatical component

The goal of this section is to describe a grammatical component to supply information to natural language interfaces with unlimited vocabulary like a text-to-speech system by rule for speech synthesis, will include the following related modules :

- a) a lexicon, where items are listed and full linguistic information is provided; lexical items are chosen so as to cover all most frequent words of the language or sublanguage in use; also words are accessed in subdictionaries according to their frequency of usage and function;
- b) a morphological analyzer composed of subdictionaries of affixes (prefixes, suffixes for inflection and derivation, stems), plus a list of lexical rules triggered by affixes to enable matching procedures to recognize the input string as a legitimate word of the language;
- c) a syntactic pre-analyzer which deals only with lexical categories resulting from previous segmentation processes and tries to strip redundant categorial information off each analysed lexical item; it disambiguates multiply tagged words by investigating its local right and left contexts and activating a transition table or deactivating it;
- d) a syntactic parser in the form of ATN (Augmented Transition Networks) which basically produces a tree structure where to higher nodes or leaves major syntactic constituent labels are associated together with grammatical functions and thematic roles.

The component is concerned with the interrelations existing between linguistic knowledge in terms of grammar – lexicon, morphology and syntax/semantics – and the fields of phonetics and phonology more closely related to speech production. Thus it is directed towards readjusting wrong decisions made at phonological level, as well as supplying the best available decision when higher information is indispensable and cannot possibly be derived only at a phonetic-phonological level.

A morphological and syntactic analyzer interacts with a speech synthesizer in particular in the following :

- a) segmentation and phonemicization procedures when diphthongs and hiatuses are relatable to the presence/absence of affixes or exceptional forms to be detected in the lexicon;
- b) words stress rules to tell homographs apart with alternate stress on different syllables of the same graphic word: this affects both syntactic and semantic analysis;
- c) homographs and homophones which lead the syntactic parser to building a final structure different from the actual one. This structure building could affect decisions made at point (d) below;
- d) by building major syntactic structures, the analyzer will enable a text-to-speech system to assign optional pauses, to let external sandhi rules apply when required but prevent them from applying when unallowed;
- e) it guides the system in assigning (and recognizing, too) the appropriate intonational structure to every utterance in accordance with meaning differences – not relatable to expressiveness, however;
- f) a differentiation between syntactically marked intonational structures which must be assigned a different intonational and pitch contour from syntactically unmarked or normal intonational structures; as well as basic stylistic phenomena quite common in Italian and French like right/left dislocations etc.;

And here is a description of the four subcomponents of our grammar, with a specification of their content.

4.1. *The Lexicon*

- from the frequency list of lexica of Italian already being produced, derive a common core of most frequent basic lemmas for lexical categories¹: Noun, Adjective, Verb;

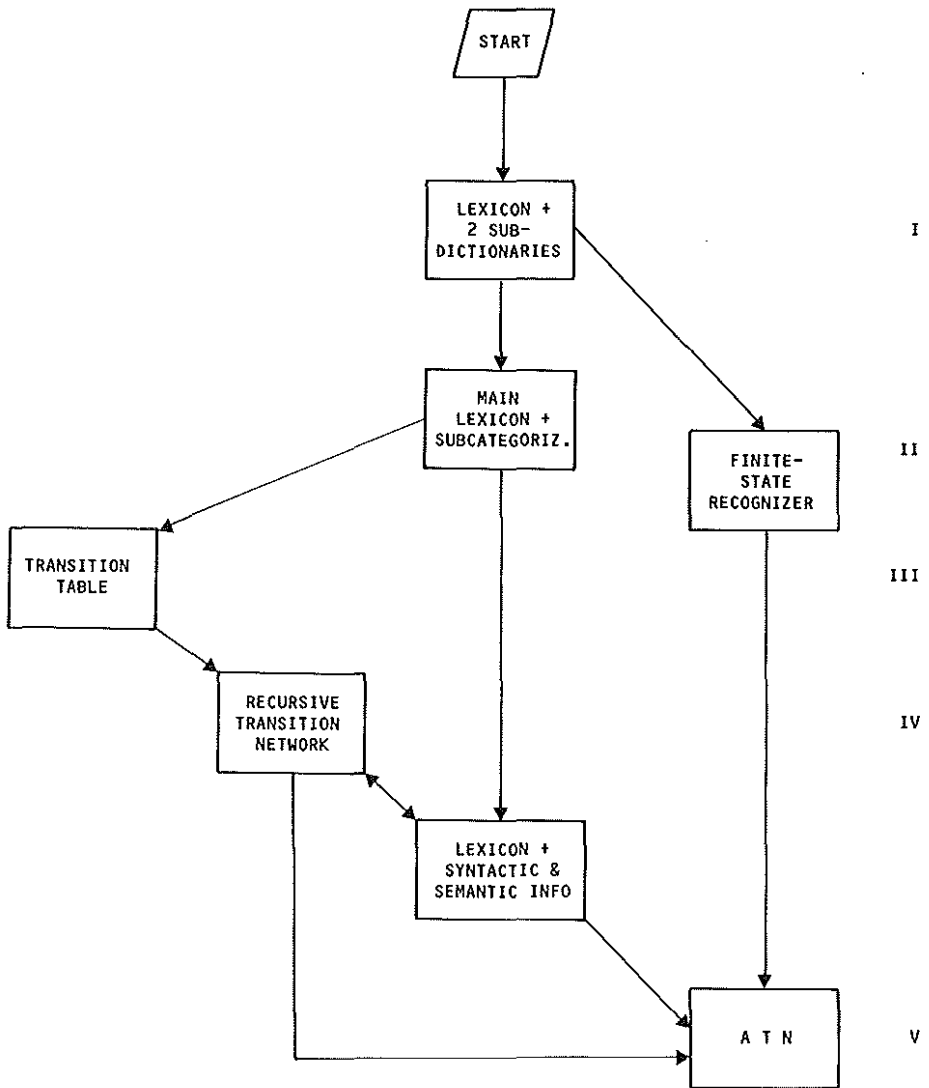


fig. 6
The grammatical component processor

- produce a comprehensive table of function or grammatical words, or closed class items which include the following : prepositions, articles, critics, personal pronouns, relative pronouns and question words, adverbs, conjunctions, possessive adjectives and pronouns, quantifiers, intensifiers, indefinite adjectives and pronouns;
- from a commonly used dictionary of Italian, extract a list of commonly used words, possibly in terms of morphological derivation, compounding and create a list of stems for economics, politics, and business application, as well as scientific and technological ones;
- information attached to each lexical item is so far made up of lexical categories or tags to which each item may belong. Syntactic and semantic information will then be added to each tag as specified below :
 - add subcategorization frames to verbs and number of controllable/governable arguments to each predicate;
 - add grammatical functions and thematic roles assumed by each argument of the predicate;
 - add semantic inherent features to nouns and adjectives, and verbs;
 - add semantic adequacy conditions for all arguments of the predicates;
 - add semantic compatibility conditions for optional arguments and adjuncts.

4.2. *The Morphological Analyzer*

It contains subdictionaries of affixes :

- to generate/recognize all inflected verb forms a tree/graph structure will be built in which nodes/leaves contain morphological information – voice, aspect, tense, person, gender – to be made available at higher constituent level;
- to generate/recognize the most frequent derivations for Noun and Adjective and Adverb by building graph/tree structures as above; for each suffix a set of phonological rules and phonotactic conditions will have to be specified to account for boundaries modification for different word types;
- in a separate tree/graph noun/adjective alterations and enclitics will be included to be segmented first, if present;
- the main lexicon will include also the most frequent stems or semi-words used to build up/recognize compound words;

- eventually, a subdictionary includes the most frequently used prefixes, which also will have to be segmented at the end, in a recursive manner – some prefixes can be present twice or three times; also in this case phonological rules will have to be incorporated for each prefix;
- when roots/themes are affected by morphological variations, allomorphs of the same basic lemma, the inflected/derived word is listed as a separate entry with appropriate information.

4.3. *The Syntactic Pre-analyzer*

A transition table has been formulated in order to disambiguate multiple tagged words, or homographs. Homographs are quite common in every language : each entry in the lexicon is assigned all category labels that it can assume according to context. Thus, inspection of the local context – right and left – is the only means available to reduce ambiguity if not to eliminate it. Disambiguation procedures are necessary to avoid useless and time-consuming backtracking steps to be taken by the parser when a multiple tagged word is encountered and multiple analyses have to be produced.

The pre-analyzer specifies for each sequence of category labels in input the allowed/forbidden transitions on the basis of major or higher constituency rules, embodied by ATN grammar and to be produced by the parser.

Thus, if lexical entries have been classified by means of a total number of twenty category labels, the preanalyzer will have to specify $20 \times 20 = 400$ steps of allowed/forbidden transitions in terms of local restrictions on the combination of category symbols. Allowable combinations will be chosen according to context-free rules specified for the ATN grammar, as well as on the basis of subcategorization frames information.

Cooccurrence conditions will eventually be formulated after an evaluation test on a sizeable sample has been carried out.

4.4. *The Parser*

Morphological features associated with each word in a text, as well as category labels filtered out by the pre-analyzer, if

required, constitutes the backbone on which the parser will work to produce its recognition procedure.

Other linguistic material, though, is available to the parser and will make the analysis more reliable and smooth: subcategorization frames are used as the most predictable paths to be followed after a major constituent has been successfully parsed. Moreover, semantic features and selection restriction rules are introduced as conditions on arcs to disentangle the parser when required. Also, morphological features can be used as conditions for tests such as agreement, which is very important and reliable for constituency in Italian.

The parser is built in the form of a top-down ATN (Augmented Transition Network), trying to restrict backtracking to ambiguous cases by means of a one word look-ahead mechanism. ATN is the best documented syntactic analyzer nowadays and the most reliable.

It includes the following components :

- a grammar of context-free rules in the form of transition networks, each network being constituted by major syntactic category symbols expanded by lexical category labels;
- the augmentation of the transition network, to be worked out on arcs and states in the form of conditions and actions.

Symbols and values generated by the prosodic translator

[1] marks primary stress in open syllable
 [2] marks secondary stress
 [3] marks unstressed syllable
 [4] marks semiconsonants [j] and [w] in falling and rising diphthongs and triphthongs
 [6] marks primary stress in closed syllable, when followed by sonorants
 [7] marks primary stress in closed syllable, when followed by consonant clusters
 [-] marks junctures, grammatical pauses or pseudo-pauses
 [+] marks non pauses or transitions between two words
 [++] marks possible conditions for phonosyntactic strengthening
 [--] marks clause boundary pause and comma intonation for parentheticals
 [=] marks utterance boundary pause and Intonational Group boundary
 [##] marks end of text pause and Intonational Group boundary
 [A] marks questions
 [B] marks exclamatives
 [I] marks imperatives

From readjustment rules and focus and saliency assignment rules

[5] marks reduced unstressed vowels, with schwa-like effect and rewrites value [3]
 [9] marks demoted stressed vowels and rewrites values [1 6 7]
 [*] marks reinforced long consonants in word body
 [0] marks elision of vowels due to assimilation at word boundary
 [^] marks potentially prominent or salient syllables
 [<] marks prominence demotion or falling pitch contour or "comma intonation"
 [⊗] marks logical focus or sentence accent
 [>] marks phonological focus or emphatic/contrastive stress

fig. 7

Example

```

6 7 5 3 5 3 1 5 3 2 3 6 3 3 3 43 17
SENTA--NON+ME+NE+IMPORTA+UN+AZZIDENTE+DI+KI+SIA+LEI=#
⊗0 >0 1 ⊗2

17 5 3 7 7 1 3 3 41 7
%AO--BELL'OMETTO--KOME+TI+KIAMI=#
⊗0 ⊗0 0 ⊗1A

7 5 7 7 2 3 2 3 6 7 3 41 3 7
+SU--SU--PRESTO--STANNO+ARRIVANDO#-XBRIGIAMOXI##
⊗0 0 ⊗01

3 3 6 5 3 6 3 7 3 46 7
HO+RAKKOLTO+UN+FUNGO+TUTTO+BIANKO##
0 1 ⊗2 ⊗3

3 3 6 3 6 3 3 1 7
STA+KANTANDO+TANTE+KANZONI##
0 1 ⊗2

3 41 5 3 3 41 3 6 7
PROVIAMO+A+KAMBIARE+KAMPO##
0 ⊗1 ⊗2

3 2 1 5 3 6 3 6 3 1 7
NON+E++STATO+UN+KOLPO+MOLTO+FORTE##
0 1 ⊗2 ⊗3

43 3 1 3 2 1 3 7 3 3 7 3 1 7
QWESTA+KORDA+E++POKO+GROSSA-E+TROPPO+KORTA##
0 1 2 ⊗3 ⊗4

3 7 3 5 3 6 3 2 3 1 5 3 1 7
HA+DEITTO-KE+INTENDE+RIMANERE+A+%ENA##
0 1 ⊗2 ⊗3

3 7 3 3 2 5 3 1 3 41 7 5 41 3 3 41 3 3 7
-IMPIXZIATI+DE#I+AFFARI+TUOI--%I+SIAMO+SKOXZIATI+DI+TE##
⊗0 1 ⊗0 ⊗1 0 ⊗1

1 7 7 3 3 43 1 3 7 3 1 7 46 3 43 6 3 23 3 6 3 3 1 3 7
KARA--LA/IATI+GWARDARE+DA+VIXINO--QWANTO+SEI+BELLA-KOI+KAPELLI+LA/ATI+JU+##
⊗0 ⊗0 1 0 ⊗1 0 1 2 ⊗3 ⊗4

```

3 6 7 1 3 3 3 3 7 7 5 2 3 3 1 7 6 3 7
 +JOVANNI--XBRIGATI+TI+HO+DETTO#+MI+SONO+XBRIGATO--KALMATI##
 @0 @0I 0 @0 @0 I

6 7 6 3 3 1 3 3 3 2 3 6 7 5 4 1 3 3 1 3 7 6 7
 +FRANKA--RENDITI+UTILE+PER+UNA+VOLTA#+%I+SIAMO+KALMATI+DA+TEMPO##
 @0 0 1 2 0 1 @2

7 3 3 7 4 3 3 1 3 7 5 4 3 3 1 7
 -XMETTILA-KON+QWELLA+MUXIKA--%I+HAI+STANKATI##
 @0 I @0 @0

3 1 3 8
 +TELEFONO##
 0@

3 2 3 7
 +NE%ESSITA+##
 0@

3 2 3 1 8
 +TELEFONATA##
 0@

3 1 0 2 6 3 2 3 1 8
 +IL+MARE+E++MOLTO+AJITATO##
 0 1> 2@

3 1 8 3 6 3 2 3 1 8
 +IL+MARE--E++MOLTO+AJITATO##
 0 1> 2@

4 3 3 2 3 1 3 2 2 3 2 3 1 3 3 6 3 3 3 2 3 1 3 2 3 2 4 3 1 3
 -QWESTA+E++LA+VO%E+DELL'ELABORATORE+DEL+%ENTRO+DI+SONO+LOJIA+KOMPUTAZIONALE+
 0 1 2 3 4

2 2 3 2 3 7 3 1 3 8
 +DELL'UNIVERSITA++DI+PADOVA##
 5 6@

4 6 3 2 3 6 8 5 3 4 1 5 3 3 1 3 8 1 3 3 4 0 3 6 8 1 3 3 1 3
 +QWANDO+SONO+PRONTA--TI+RIKIAMO+AL+TELEFONO--SOLO+SE+HAI+I+SOLDI--FAI+TU+UNA+
 0 1@ 2 3 4 5 0 1

3 2 3 1 3 3 1 8
 +TELEFONATA+DOMANI##
 2@ 3@

3 3 1 5 3 7 8 7 5 3 6 3 8 5 7 4 3 7 3 3 3 1 8 5
 +A+LI>ANO+IN+LU<0--JA++ALLE+UNDI%I--LA+SABBIA-SKOTTA+PER+IL+SOLE--E+
 0 1 2 3@ 0 1 2@

3 1 3 3 1 3 3 6 3 2 3 3 2 3 7
 +BA>ARSI+NEL+MARE+DIVENTA+UNA+NE%ESSITA+##
 0 1 2 3@

3 1 3 3 2 3 1 3 3 3 3 7 8
 -VORREI-KE+FOSSE+KOME+SE+TU+%I+FOSSI##
 0 1 2@

3 2 3 7 3 3 3 6 3
 +MI+HANNO+DETTO-KE+%E+L'HANNO##
 0 1@

3 3 1 3 3 3 1 3 3 1 8
 +TI+RIPETO-KE+DOMANI+%I+SONO##
 0 1@ 2@

7*3 7 43 3 1 3 3 23 3 7 5 7 3 7
 VO<O+PROPRIO+VEDERE+KI+RIU/IRA++A+BATTERMI=#
 0 1 2 3 @4 B

3 7*3 6 3 1 7 1 3 1 3 3 2 6 3 1 7
 TI+VO<O+TANTO+BENE--T'AMO+KOME+NESSUN'ALTRA+MAI=#
 0 @1 @2 0 1 @2 @3

3 1 5 7 2 3 7 43 7 3 6 3 7 3 1 7
 OH+NO--MA+NON+HANNO+PROPRIO-NESSUN+ALTRO+DA+SKO%ARE-----+
 >0 1 2 @3

7 1 5 3 3 7 3 6 7 5 3 1 7
 NO--NO--NON+<I+HO+DETTO+NULLA#+TE+LO+JURO=#
 >0 1 @0B

3 3 1 3 2 3 7 3 7 7 3 3 1 3 7
 O+<I+SKRIVI+UNA+LETTERA--O+<I+TELEFONI##
 0 @1 @0

3 1 3 3 1 6 3 7 3 3 4 1 3 3 3 1 3 3 7 3 7
 EGREJO+SI>OR+SINDAKO--SONO+LIETO+DI+POTERLE+PROMETTERE-----
 0 @2 0 1 @2

3 1 3 3 3 1 7 7 5 3 1 5 3 1 3 2 3 1 3 7
 +SI>ORE-E+SI>ORI--EKKO+A+VOI+UN+DUO+FORMIDABILE=#
 0 @1 <0 1 2 @3 B

2 3 1 3 3 7 1 0 3 1 5 4 7 2 4 3 1 7
 UNA+KOXA+KOXI++-KARA+A+TE--NON+PUO++DISPIA%ERMI##
 0 1 @2 >0 @1

2 3 1 3 14 3 3 7 3 1 3 3 7 43 13 3 6 3 7 7
 PRENDEREMO+L'AUTOBUS--O+IL+FILOBUS--O+QWALSIAXI+ALTRO+ME&&O##
 0 @1 @0 <0 <1 @2

4 3 3 3 7 3 7 3 1 3 1 7
 QWESTA+MINESTRA+FA+DAVVERO+SKIFO=##
 0 @1 @2B

7 5 2 1 7 3 3 3 7 3 7 1 5
 NO--NON+E++VERO--SONO+KALMISSIMO--IO##
 >0 @0 @0

3 1 3 7 3 7 43 4 7 3 1 3 3 6 7
 ORMAI+NON+HO+PROPRIO+PIU++NESSUNA+SPERANZA-----
 0 >1 2 >3 >4

3 1 3 3 7 3 43 7 3 7 1 7 5 1 7
 EXIGO-KE+TUTTO+SIA+FATTO+PER+BENE#-KAPITO=#
 0 1 2 @3 @0A

3 1 3 1 3 7 5*1 3 0 3 1 7 3 3 1 3 3 7
 U/ITE+SUBITO=-LA/ATEMI+IN+PAXE=-ASPETTATE+DI+LA+=#
 ^ ^ B ^ I ^ B ^ ^ B

3 3 7 1 3 7 5 2 3 7 1 3 7
 +LO+PORTO++IKARO--LO+KONZEPI++IKARO##
 ^ ^ ^ ^

3 2 3 7 5 3 1 3 7 3 3 3 7 5 3 1 3 7 3 5 3 1 3 7
 +LO+DEFINI+INAJIBILE--UNA+%ITTA++INAJIBILE--E++INKREDIBILE=#
 ^ ^ ^ ^ ^ ^ ^ ^ ^ ^ ^

2 3 3 7 5 1 7 3 3 3 7 5 7 7 5 3 7 5 4 7 7
 +UNA+%ITTA++APERTA--UNA+%ITTA++INTATTA--LO+PORTO++INDIETRO=#
 ^ ^ ^ ^ ^ ^ ^ ^ ^ ^ ^

3 3 1 3 2 3 1 3 7 5 4 7 3 2 3 7 7
 +UN+PROBLEMA+INSOLUBILE--UN+PIATTO+INDIESTO##
 ^ ^ ^ ^ ^ ^ ^ ^ ^

3 3 4 1 3 7 3 7 3 3 3 6 3 1 3 7
 +UN'UNIONE+IBRIDA--UNA+STUPENDA+ISOLA##
 ^ ^ ^ ^ ^ ^ ^ ^ ^

2 3 2 3 2 3 4 1 5 3 1 3 7 3 3 3 4 1 5 3 1 3 7 7 3 5 3 1 7
 +UNA+RAPPRESENTAZIONE+IKONIKA--UNA+FUNZIONE+IKONIKA--LA/IAMI+ANDARE--
 ^ ^ ^ ^ ^ ^ ^ ^ ^ ^ ^ ^ ^

4 6 3 1 5 3 1 7 5 1 7 1 7
 QWANDO-TORNA+IL+PAPA=-DOMANI--KREDO##
 ^ ^ ^ ^ ^ ^ ^ ^ ^

4 1 0 3 4 7 7 5 3 2 3 1 7 5 7
 VIENI+IN+VIAJJO=+TI+DIVERTIRAI--PAPA+=##
 ^ ^ ^ ^ ^ ^ ^ ^ ^

1 3 3 3 7 5 7 7*3 3 1 3 4 7
 ABITO+LAJJU+#+E+NON+VO<O+VERDERTI+PIU+=#
 ^ ^ ^ ^ ^ ^ ^ ^ ^

3 6 0 3 7 3 7 1 7 5 1 3 3 2 3 7 5
 AVANTI+IL+PROSSIMO--NOME--KO>OME-E+INDIRIZZO##
 ^ ^ ^ ^ ^ ^ ^ ^ ^

4 3 3 3 7 3 7 5 3 1 7 2 3 1 3 7 5 3 1 3 7
 QWESTO+DIRATTITO----HA+LO+SKOPO----PROGRAMMATIKO----DI+DISKUTERE----
 ^ ^ ^ ^ ^ ^ ^ ^ ^

3 2 3 1 3 7 3 3 1 3 7 3 3 7 3 7
 LA+PROBLEMATIKA----SINTOMATIKA----FUTURISTIKA##
 ^ ^ ^ ^ ^ ^ ^ ^ ^

5 3 3 2 3 1 3 7 2 3 6 3 7 4 7 3 3 7 7 3 3 4 1 7
 #-TRA+I+LAVORATORI+%E+++ERTAMENTE+RABBIA--AMAREZZA--DELUXIONE##
 0 1 2 @3 @0 @0

2 3 7 3 3 7 1 3 3 4 1 5 3 3 7 3 2 3 3 6 7 5 3 1 4 7
 UNA+LOTTA+KOXI++DURA+SI+KIUDE+AL+DI+SOTTO+DELLE+SPERANZE--E+ANKOR+PIU++
 0 1 2 3 4 @5 <0 1

6 3 2 3 4 1 3 3 7 3 3 1 3 2 3 3 3 3 5 3 1 3 2 3 2 3 1 3
 DELLE-ILLUXIONI-KE+ESSA-AVEVA+SU/ITATO+(E-KE+ALKUNI+HANNO+E%ITATO+
 2 3 4 5 @0 0 1

2 3 6 5 3 3 2 3 1 5 3 1 3 3 2 3 4 1 3 4 3 4 7 3 2 3 1 3
 KOMMETTENDO+UN+PERIKOLOXO+ERRORE)#-L'ESTROMISSIONE-DEI+QWADRI+SINDAKALI+
 2 @3 @4 0 1 2

2 3 1 3 3 7 5 3 2 3 4 1 3 7 3 1 3 3 3 1 3 7 3 1 3 3 1 3
 ATTRAVERSO+LA+KASSA+INTEGRAZIONE+FA+TEMERE-KE+LA+FIAT-POSSA+ORA+PASSARE+
 3 4 5 0 1 2 3 4

3 3 2 3 4 1 7 5 3 3 7 3 3 6 3 3 2 3 7 3 2 3 6 3 3
 KON+LA+REPRESSIONE#-E+LA+KAMPA>A+DI+STAMPA+HA+INTRODOTTO+ELEMENTI+DI+
 @5 <0 1 2 3

2 3 4 1 7 3 3 4 3 3 6 7 6 5 3 1 3 3 7 3 3 3 2 3 1 3
 KONFUSIONE--DIXORIENTAMENTO--MENTRE+I+SOLITI-GRUPPI+DI+PROVOKATORI+
 @4 @0 0 1 2 3

1 3 3 0 3 2 3 1 5 3 2 3 6 3 2 3 3 2 3 1 7 3 3 6 3
 %ERKANO+DI+INKANALARE+IL+MALKONTENTO+KONTRO+IL+SINDAKATO--INNALZANDO+
 4 5 0 @1 0

3 3 4 1 3 2 2 3 7 7 5 3 3 2 7 3 6 5 3 2 3 4 1 3 3 1 3
 LA+BANDIERA+DELL'OLTRANZIXMO--MA+IN+REALTA++RENDENDO+UN+ULTERIORE+SERVIJO+
 1 @2 0 1 2 3

3 2 3 1 7 7 4 3 3 7 4 7 5 2 3 6 3 3 6 3 3 6 3 3 1 3
 AL+PADRONATO#>DA+QWELLA+RABBIA--DA+SENTIMENTI+PROFONDI-KE+VANNO+KAPITI+
 @4 @0 0 1 2 3

3 2 3 1 7 5 1 5 3 1 3 3 1 3 3 3 4 1 3 2 3 2 3 1 3 3 4 1 3
 E+RISPETTATI--OKKORRE+INVE%E- PARTIRE+PER+KOSTRUIRE+UNA+KOMBATTIVA+JESTIONE+
 @4 0 1 2 3 4 5

3 3 1 7 7 7 1 3 3 1 3 2 2 3 3 1 3 7 7 5 3 1 3
 DELL'AKKORDO#>NON+%E++SOLO+LA+KRIXI+DELL'AUTOMOBILE#-%'E++UNA+KRIXI+
 @0 >0 >1 2 3 0 1

3 3 2 3 7 3 2 3 4 1 4 3 2 3 3 7*0 3 6 7 5 3 4 1 3 3 3 2 3 1 3
 E+DEL+DIXIMPE>O-FINANZIARIO+DELLA+FAMI<A+A>ELLI#>LA+JESTIONE-E+LA+STRATEJIA+
 0 1 @2 @3 0 <1

2 5 3 1 3 3 1 2 6 1 5 3 7 0 3 1 4 3 3* 3 1 7 0 2
 +SONO+I+VERI-PROBLEMI+(NON+%ERTO+IL+KOSTO+ORARIO+DEL+LAVORO--KE+E++
 2 @3 >0 >1 2 3 @4

2 3 6 5 3 3 7 3 3 4 3 3 2 5 3 7 4 3 2 3 6 3 5 0 3 7 3
 NETTAMENTE+AL+DI+SOTTO+DI+QWELLO+DELLE+INDUSTRIE+KONKORRENTI)#>SE+SI+IMBOKKA+
 0 1 2 @3 0

3 1 3 3 3 2 3 6 3 3 7 7 5 7 3 3 3 6 3 2 4 3 1 3
 LA+STRADA+DEL+RIXANAMENTO+DEL+GRUPPO--BEN+PRESTO+SI+DOVRANNO+RIKIAMARE+
 1 2 @3 0 1 2

3 7 3 3 6 3 3 2 3 1 3 7 0 3 7 5 3 2 3 4 1 7 5 4 3 3 7
 IN+FABBRIKA+MOLTI+LAVORATORI-OJJI+IN+KASSA+INTEGRAZIONE#>SE+QWESTO+NON+
 3 4 5 0 1 @2

3 1 7 5 7 3 2 3 3 1 3 7 5 7 3 4 1 5 3 1 3 2 3 2 3 7 7
 +SU%EDE--SI+APRE+UNA+VORAJINE#>DI+%O++DOBBIAMO+AVERE+KONSAPEVOLEZZA##
 @0 0 1 0 1 2 @3

3 6 3 1 3 3 7 7 6 4 7 5 7 3 4 3 3 2 4 3 6 3 2 3 3 4 1 3
 LO-SKONTRO+ORA+SI+SPOSTA--DUNGWE--DAL+BLOKKO+DEI+LI%ENZIAMENTI+ALLA+JESTIONE+
 0 1 @2 @0 0 1 2

3 3 1 3 5 3 3 2 3 6 3 2 3 1 3 5 2 5 3 4 1 3 3 7*3
 DELL'AKKORDO-E+AL+RIXANAMENTO+DELLA+FIAT#>ED+E++UNA+NUOVA+BATTA<A
 @3 <0 1 0 1

2 3 7 5 3 3 1 4 2 5 2 3 6 3 2 2 3 7 3 3 6 3 2 3
 DELLA+KLASSE+OPERAIA--KE+KORRISPONDE-ALL'INTERESSE+PROFONDO+DELLA+
 2 @3 0 1 2

3 2 3 7 2 4 3 1 7 5 4 3 3 2 3 7 6 3 3 3 2 3*1 7 6 0 2 3 4 1 7
 KOMUNITA++NAZIONALE#SE+QWESTE+VERITA++VENGONO+RIKONO/UTE--MOLTI+ITALIANI--
 3 @4 0 1 @2 0 >1

6 3 3 1 3 1 3 7 5 6 3 1 3 1 4 3 3 4 1 3 3 2 3 1 3 2 3 1 3
 ANKE+DI+%ETI+DIVERSI--DOVRANNO-DIRE+"GRAZIE"+A+QWEI-LAVORATORI+DELLA+FIAT+
 0 1 @2 0 1 2 3 4 5

3 6 3 1 3 3 3 2 3 1 3 3 3 3 1 3 7 3 3 1 3 3 3 3 6 4 7
 KE+ALTRI+%ERKANO+DI+DIFFAMARE-E+DI+SPOKARE-KON+LE+BUJIE-E+LE+KALUNNIE
 0 1 2 <3 @4 @0

7 3 3 1 3 7 1 3 3 3 7 5 2 3 1 7
 METTILO+PURE+LI++POI+<I+DARO++UN'OKKIATA##
 @0 I 0 1 2 3 @4

7 7 2 3 1 5 3 3 2 3 1 7 5 1 3
 +SI++-SI++-E++DAVVERO+UN+BEL+KAPPELLINO--SI>ORA##
 0 1 2 @3 @0

4 3 3 1 5 3 3 3 3 1 7 7 7 3 1 7
 SEI+DAVVERO+UN+BEL+KAMPIONE=+NON+%E++KE+DIRE=#
 0 @1B >0 >1B

3 7 3 4 3 3 1 7 1 3 3 7 1 7
 E+TUTTA+QWESTA+ROBA--KOXA+SARA++MAI=#
 0 @1 0 @1 @2

3 7 3 7 3 3 7 4 3 4 6 7 4 3 3 3 1 7
 MA+NON+KAPISKONO+PROPRIO+NIENTE--QWESTI+KRETINI=#
 >0 1 @2 @0B

7 5 6 5 3 7 1 3 3 1 3 7 3 1 7
 STA++ATTENTO+A+NON+FARTI-VEDERE+DA+NESSUNO=#
 0 1 >2 >3 @4B

1 7 3 1 3 7 7 3 1 5 3 1 3 4 7
 PREGO--S'AKKOMODI--FA%%A-KOME+A+KAXA+SUA=#
 @0 @0 0 1 @2 B

3 2 1 3 3 1 7 7 7 3 1 7
 NON+E++MIKA+KADUTA--S'E++FATTA+MALE=#
 @0 @1A 0 @1 @2A

1 4 3 3 1 0 2 3 1 3 4 3 3 7 3 1 3 7
 +E+IO+DOVREI+INDOSSARE-QWESTO+STRA%%O+D'ABITO=#
 0 1 2 3 @4 A

3 7 3 1 7 3 2 3 1 3 7 3 3 7 5 7 7
 NON+S0-KE+DIRE--E++KAPITATO+TUTTO-KOXI++IN+FRETTA=#
 @0 0 1 2 @3 B

4 3 4 1 7 5 1 7 5 7 3 2 6 3 2 3 1 3 3 6 7
 +SIA+BUONO--LA+PREGO++KON%EDA-UN'ALTRA+SETTIMANA+DI+TEMPO##
 @0 @0 0 1 2 @3

3 4 6 3 6 5 3 1 3 3 3 3 7 3 1 3 7
 MA-QWANTE+VOLTE+ANKORA+TE+LO+DOVRO++RIPETERE=#
 0 1 2 @3 @4 B

2 3 2 3 1 7 5 3 1 3 2 3 4 1 3 3 6 3 7 5 1 7
 +SONO+KOSTERNATO--NON+AVEVO+L'INTENZIONE+DI+ROMPERLO#MI+SKUXI##
 @0 >0 1 2 @0

Bibliography

- Antinucci, F. & Cinque, G. (1977). Sull'ordine delle parole in italiano : l'emarginazione, *Studi di Grammatica Italiana*, Vol. VI, Firenze, 121-146.
- Bates, M. (1978). The Theory and Practice of Augmented Transition Network Grammar, in L. Bole (ed.), *Natural Language Communication with Computers*, Springer, New York, 191-260.
- Berman, A. & Szamosi, M. (1972). Observations on Sentential Stress, *Language*, 8, 2, 304-325.
- Bertinetto, P.M. (1981). *Strutture Prosodiche dell'italiano*, Accademia della Crusca, Firenze.
- Bolinger, D. (1973). Accent is Predictable - If you're a Mind-Reader, *Language*, 48, 3, 633-645.
- Boyce, S. & Menn, L. (1979). Peaks Vary and Endpoints Don't : Implications for Intonation Theory, *Proceedings of the Fifth Berkeley Linguistic Society Meeting* (Berkeley Linguistic Society, Berkeley).
- Bresnan, J. (1971). Sentence Stress and Syntactic Transformations, in Hintikka et al. (eds, 1973), *Approaches to Natural Language*, Reidel, Dordrecht, 3-47.
- Bresnan, J. (1972). Stress and Syntax : a Reply, *Language*, 48, 2, 326-341.
- Bresnan, J. (1978). A Realistic Transformational Grammar, in M. Halle et al. (eds), *Linguistic Theory and Psychological Reality*, MIT Press, Cambridge Mass., 1-59.
- Bresnan, J. (1980). Polyadicity : Part I of a Theory of Lexical Rules and Representations, in Hoekstra et al., *Lexical Grammar*, Foris, Dordrecht.
- Bresnan, J. (1981). Control and Complementation, *Linguistic Inquiry*, 13, 3, 343-434.
- Bresnan, J. (1982). *The Mental Representation of Grammatical Relations*, MIT Press, Cambridge Mass.
- Brown, G. et al. (1980). *Questions of Intonation*, Croom Helm, London.
- Canepari, L. (1980). *Italiano Standard e Pronunce Regionali*, CLEUP, Padova.

- Cappelli, A., Ferrari, G., Moretti, L., Prodanof, I. & Stock, O. (1978). Parsing an Italian Text with an ATN parser, Laboratorio di Linguistica Computazionale, C.N.R., Pisa.
- Chomsky, N. (1970). *Deep Structure, Surface Structure, and Semantic Interpretation*, Mouton, The Hague.
- Chomsky, N. (1973). *Conditions on Transformations*, tr. it. *Forma e Interpretazione* (1980), Il Saggiatore Milano, pp. 153-251.
- Chomsky, N. (1975). Conditions on Rules of Grammar, *Linguistic Analysis*, 2, pp. 303-352.
- Chomsky, N. (1977). On WH-Movement, in Culicover et al. (eds), *Formal Syntax*, Academic Press, New York, 71-132.
- Chomsky, N. (1981). Principles and Parameters in Syntactic Theory, in N. Hornstein, D. Lightfoot (eds), *Explanations in Linguistics*, Longman, New York.
- Chomsky N. & Halle M. (1968). *The Sound Pattern of English*, Harper & Row, New York.
- Cinque, G. (1972). Grammatica Generative e Metafora, *Studi di Grammatica Italiana*, Vol. II, 261-295.
- Cinque, G. (1974). "Presupposizioni" di voci lessicali e di costrutti e loro rilevanza sintattica, in AA.VV., *Fenomeni Morfologici e Sintattici dell'Italiano Contemporaneo*, S.L.I., Vol. 1, Bulzoni, Roma, 461-469.
- Cinque, G. [a cura di] (1979). *La Semantica Generativa*, Boringhieri, Torino.
- Cinque, G. (1984). Clitic Left Dislocation in Italian and the Move-Parameter, (ms.) Università di Venezia.
- Cooper, W.E. (1980). Syntactic-to-Phonetic Coding, in *Language Production*, B. Butterworth (ed.), Academic Press, New York, 296-333.
- Cooper, W.E. & Paccia-Cooper, J. (1980). *Syntax and Speech*, Harvard U.P., Cambridge Mass.
- Cordin, P. (1981). Ipotesi Generative e Condizioni per l'interpretazione Anaforica dei pronomi personali, su *Atti del Seminario "Sull'anafora"*, Accademia della Crusca, Firenze, 69-81.
- Delmonte, R. (1981). "L'accento di parola nella prosodia dell'enunciato dell'Italiano standard", *Studi di Grammatica Italiana*, Accademia della Crusca, Firenze, 351-394.

- Delmonte, R. (1983a). A Phonological Processor for Italian, *Proceedings of the First Conference of the European Chapter of the Association for Computational Linguistics*, Pisa, 26-34.
- Delmonte, R. (1983b). *Regole di Assegnazione del Fuoco o Centro Intonativo in Italiano Standard*, CLESP, Padova.
- Delmonte, R. [a cura di] (1983). *Linguistica computazionale*, CLESP, Padova.
- Delmonte, R. (1984). "La Syntactic Closure nella Teoria della Performance", *Quaderni Patavini di Linguistica*, 4, CLESP, Padova, 101-131.
- Delmonte, R., Mian, A. & Tisato, G. (1984). A text-to-Speech System for the Synthesis of Italian, *Proceedings of ICASSP*, San Diego (California).
- Emonds, J.E. (1976). *A Transformational Approach to English Syntax*, Academic Press, New York.
- Fillmore, C. (1969). Types of Lexical Information, in F. Kiefer (ed.), *Studies in Syntax and Semantics*, Reidel Dordrecht, 109-137.
- Gazdar, G. (1982). Phrase Structure Grammar, in P. Jacobson & G.K. Pullum (eds), *The Nature of Syntactic Representation*, Reidel, Dordrecht, 131-186.
- Gueron, J. (1980). On the Syntax and Semantics of PP Extraposition, *Linguistic Inquiry*, 11, 4, 637-677.
- Gussman, E. (1980). *Studies in Abstract Phonology*, Linguistic Inquiry Monograph Four.
- Hirst, D. (1977). *Intonative Features*, Mouton, The Hague.
- Hoekstra, T., Hulst & Morrtgaat, eds (1980). *Lexical Grammar*, Foris, Dordrecht.
- Hulst, H. & Smith, N., eds (1982). *The Structure of Phonological Representations 1*, Foris, Dordrecht.
- Jackendoff, R. (1972). *Semantic Interpretation in Generative Grammar*, MIT Press, Cambridge Mass.
- Jacobson, P. & Pullum, G.K., eds (1982). *The Nature of Syntactic Representation*, Reidel, Dordrecht.
- Kaplan, R.M. (1973). A general syntactic processor, in R. Rustin (ed.), *Natural Language Processing*, Algorithmic Press, New York, 193-241.

- Kay, M. (1977). Morphological and syntactic analysis, in A. Zampolli (ed.), *Linguistic Structure Processing*, North Holland, New York, 131-234.
- Kay, M. (1979). Functional Grammar, *Proceedings 5th Annual Meeting of the Berkeley : ling. Society*, 142-158.
- Kuroda, S.Y. (1969). Remarks on Presuppositions and Selectional Restrictions, in F. Kiefer (ed.), *Studies in Syntax and Semantics*, Reidel, Dordrecht, 138-167.
- Ladd, D.R. (1978). *The Structure of Intonational Meaning*, Indiana U.P., Bloomington.
- Lakoff, G. (1972a). Linguistics and Natural Logic, in Davidson & Harman (eds), *Semantics of Natural Language*, Reidel, Dordrecht, 545-664.
- Lakoff, G. (1972b). The Global Nature of the Nuclear Stress Rule, *Language*, 48, 2, 285-303.
- Lieberman, M. & Prince, A. (1977). On Stress and Linguistic Rhythm, *Linguistic Inquiry*, 8, 249-336.
- Lieberman, M. & Sag, I. (1974). Prosodic Form and Discourse Function, in *Papers from the 10th Regional Meeting of the Chicago Linguistic Society*, Chicago, CLS, 416-427.
- Nespor, M. & Vogel, I. (1982). Prosodic Domain of External Sandhi Rules, in Hulst & Smith, *The Structure cit.*, 225-257.
- O'Shaughnessy, D. & Allen, J. (1983). Linguistic Modality Effects on Fundamental Frequency in Speech, *JASA*, 74 (4), October, 1155-1171.
- Pereira, F. (1983). Logic for Natural Language Analysis, Tech. Note 275, SRI International, Menlo Park (Cal.).
- Perlmutter, D.M. (1979). Working 1s and Inversion in Italian, Japanese and Quechua, in *Proceedings 5th Annual Meeting of the Berkeley Ling. Society*, 277-325.
- Pierrehumbert, J. (1981). Synthesizing Intonation, *JASA*, 70 (4), 985-995.
- Pinker, S. (1972). A Theory of the Acquisition of Lexical Interpretive Grammars, in Bresnan (ed.) *cit.*, Chapter 10.
- Quirk, Greenbaum & Leech (1976). *A Grammar of Contemporary English*, Longman, London.
- Radford, A. (1977). *Italian Syntax*, Cambridge U.P., Cambridge (UK).

- Ritchie, G. (1980). *Computational Grammar*, The Harvester Press, Sussex (UK).
- Rizzi, L. (1978). A Restructuring Rule in Italian Syntax, in S.J. Keyser (ed.), *Recent Transformational Studies in European languages*, MIT Press, Cambridge Mass., 113-158.
- Rizzi, L. (1980). Violations of the Wh- Island Constraint in Italian and the Subjacency Conditions, *Journal of Italian Linguistics*, 5, 1/2, 157-195.
- Rizzi, L. (1982). *Issues in Italian Syntax*, Foris Publications, Dordrecht.
- Schmerling, S.F. (1974). A Re-Examination of "Normal Stress", *Language*, 50, 1, 66-73.
- Selkirk, E.O. (1980). The Role of Prosodic Categories in English Word Stress, *Linguistic Inquiry*, 11, 3, 563-605.
- Sgall, P., Hajicova, E. & Benesova, E. (1973). *Topic, Focus and Generative Semantics*, Scriptor, Kronberg.
- Sorensen, J.M. & Cooper, W.E. (1980). Syntactic Coding of Fundamental Frequency in Speech Production, in R.A. Cole (ed.), *Perception and Production of Fluent Speech*, Lawrence Erlbaum Ass. Inc., New Jersey, 399-440.
- Stock, O. (1982). Parsing on Wednesday, Rapporto Tecnico n° 312, Istituto di Psicologia del C.N.R., Roma.
- Stockwell, R.P. (1972). The Role of Intonation, in D. Bolinger (ed.), *Intonation*, Penguin Books, 87-109.
- Stockwell, R.P., Schachter, P. & Hall Partee B. (1973). *The Major Syntactic Structures of English*, Holt Rinehart & Winston, New York.
- Wanner, E. & Maratsos, M. (1978). An ATN Approach to Comprehension, in M. Halle, J. Bresnan, G. Miller (eds), *Linguistic Theory and Psychological Reality*, MIT Press, Cambridge, Mass., 119-161.
- Winograd, T. (1983). *Language as a Cognitive Process; Vol. 1: Syntax*, Addison Wesley, Menlo Park (Cal).
- Woolford, E. (1983). Bilingual Code-Switching and Syntactic Theory, *Linguistic Inquiry*, 14, 3, 520-536.