



CONTEXTUAL SYNTACTIC ANALYSIS FOR TEXT-TO-SPEECH CONVERSION

S.Quazza*, E.Vivalda*

ABSTRACT

A grammatical analyzer has been implemented to the specific end of providing the syntactic information required, in phonetic transcription and prosody generation, by a high-quality text-to-speech system performing real-time speech synthesis from unrestricted Italian text, relying on a small dictionary. After detecting the correspondencies between syntax and phonology, contextual rules have been stated, supported with statistical analysis of written texts, predicting the syntactic role of content words on the basis of adjacent function words. Two sample applications of the grammar are reported: the solution of stress ambiguities caused by non-homophone homographs and the grouping of words into "phonological words".

INTRODUCTION

A text-to-speech system converting any text into the corresponding audible speech, must have access to knowledge concerning different levels of the language structure: as the graphemic system is not usually isomorphic with the phonetic system and only roughly represents those suprasegmental phenomena (e.g. intonation) which are linguistically relevant in spoken language, rules directly converting each phoneme into its acoustic realization are not sufficient to obtain plausible speech from written texts; rather, phonology, syntax (and semantics) should be appealed to.

In the following, a grammar codifying syntactic aspects of Italian relevant to speech synthesis is described. As no formal rule system can give a complete description of a natural language, the effectiveness of the grammar is to be measured on its peculiar aims: analysing sentences in order to collect cues to their phonic realization. The design of the grammar, tailored on text-to-speech demands and satisfying the requirements of an actual system (concerning memory space, real-time performance), relies on analyses of human utterances and of written texts finding out which syntactic features can provide the maximum specific (phonological) information with the least effort (ref 1): "function words" (articles, prepositions, pronouns, conjunctions, auxiliaries, etc.), connecting "content words" (nouns, adjectives, verbs) by specifying their syntactic/semantic function in the context, result to have a direct prosodic relevance and a key-role in detecting the syntactic structure of sentences.

RELATIONS BETWEEN SYNTAX AND PHONOLOGY IN ITALIAN

The need for higher-level linguistic information in the phonological processing of an Italian text arises in some cases in grapheme-to-phoneme conversion: although direct phonetic transcription, required by unlimited vocabulary systems, is quite simple in Italian, the phonetic context leaves open ambiguities which can be solved by morpho-syntactic rules (beside a necessary exception list).

Syntactic analysis is more severely required in the generation of prosodic parameters, that is values of duration and f_0 realizing word stress and sentence intonation.

While prosody can express emotive and pragmatic attitudes in spontaneous speech production its more linguistic aspects are highly conventional. A "prosodic structure" can be recognized in utterances, concurring to their logical structuring and affecting intelligibility, beside naturalness. The written text codifies explicitly very little prosodic information, by the insertion of punctuation marks: nevertheless, the prosodic realizations of the same text by different readers are similar, the semantic and syntactic structures of the sentence being mirrored by a 'neutral' prosodic structure. As a text-to-speech system should 'read aloud' written sentences, its prosodic performance is quite improved when it is able to detect syntactic patterns and translate them into prosodic ones.

Global prosodic phenomena -as intonation contours- corresponding to high-level syntactic features are imposed on an underlying structure determined by the hierarchical realization of lexical stresses in terms of prominence, which in turn is largely dependent on local syntactic relations. Lexical words result to be grouped together into "phonological words" uttered with a main accent and no inner word-boundary marking: only one of the lexical stresses is fully realized (Nuclear Stress Rule) (ref 2). Phonological words are delimited by boundary phenomena (such as vowel lengthening) and connected into "intonation groups" (which in turn may be bounded by pauses) marked by different intonation contours depending on their (semantic-)syntactic function in the sentence.

Examples of the syntax-prosody interactions pertaining to different levels in the prosodic structure of Italian are reported in the following.

Lexical stress location- The position of lexical stress in Italian is not determinable by

*Olivetti Corporate Research, Voice Processing Lab., C.so Svizzera 185, 10149 Torino, Italy

means of mere phonological rules. Lacking a dictionary that would provide the stress position for each word, a morpho-syntactic approach proves to be helpful. The stress behaviour shows statistical regularities; the lexicon can be classified by word-endings: the most frequent stress pattern of words in a given class is stated as default choice, while relevant exceptions are listed in the dictionary. Where different stress positions are equiprobable for words sharing a given ending, syntactic information can provide a more powerful choice criterion. For instance, words ending in "-ano" carry the stress on the penultimate syllable if nouns or adjectives, on a previous one if verbs: "roma'no" (roman), "capita'no" (captain) vs. "re'mano" (they rowed), "ca'pitano" (they happen). The distinctive role of lexical stress in Italian is proved by the existence of hundreds of minimal pairs like <capita'no,ca'pitano>, <a'ncora,anco'ra> (anchor / still,yet). Contextual information is always needed to solve the ambiguities caused by non-homophone homographs: while there are situations in which semantic and pragmatic considerations are necessary, detecting the syntactic role of the word often provides the desired solution: "ancora" will be realized as "a'ncora" if a noun is suitable to the context, as "anco'ra" if an adverb is to be preferred.

Phonological words- Factors affecting the degree of prominence to be assigned to a stressed syllable in Italian are the following:

- the position of the word in the syntagm: the inner cohesion of the syntagm is realized by a greater prominence of the tonic syllable of the last word and by boundary phenomena marking the end of the constituent. The grammatical category of the word determines its accentual realization in an indirect way, by determining the phrase structuring of the sentence (ref 3); an adjective may precede or follow the noun in a noun phrase: in the first case it will lose prominence, in the second it will carry the main stress.
- lexical peculiarities: the lesser is the informative content of a word, that is the more predictable it is in a given context, the more likely it is to lose prominence: it turns out that lexical factors directly affect prominence, "worn out" words (most probable words) being most exposed to de-accentuation (ref 4). Function words, when acting as grammatical supports of content words, introduce the syntagm and don't carry semantic information: as a consequence they are prosodically joined to the following words.
- rhythmical factors : distance between stresses, length of the word.
- global syntactic features of the sentence: for instance, interrogative adjectives are accented in questions and de-accented in statements. The length of the following syntagm may determine whether the words of a constituent should form a single phonological word or not: while demonstrative adjectives are usually linked to the following noun, they are accented if the noun is the last word in the sentence.

Pauses and intonation contours- "Intonation group" boundaries may be marked by pauses, joining a syntactic-semantic function to their physiological role. The length of a breath group may depend on the speech rate; the deeper is the syntactic boundary, the more probable is that it will be realized by a pause. Intonation contours mark the logical relations between phrases as well as the clause type or sentence modality. For example, in a sentence of standard structure subject-predicate a pause may occur at the beginning of the predicate: in any case a typical intonation contour will mark the syntagm boundary ; a parenthetical clause is marked by peculiar pitch and speech-rate; a yes-no question is realized with higher pitch, and fall-rise contour with lower speech-rate on the last word. Even at this higher syntactic level, function words have a direct prosodic relevance beside their phrase structuring role: interrogative particles, conjunctions and adverbs, together with punctuation marks, are crucial in marking sentence modality and the role of clauses.

THE GRAMMAR

The aim of the grammar here described is to determine the syntactic role of words where it can give hints about the phonetic or prosodic realization of single words or of the whole sentence. As the dictionary size must be kept to a minimum, the main concern will be, beside the choice of the correct alternative for grammatically ambiguous words, the prediction of the category of "unknown" words on the basis of the syntactic function of adjacent words, when morphological analysis fails. The Italian language does not impose rigid patterns on word sequencing. Nevertheless, function words have a definite position in the sentence, linked to that of the word, syntagm or clause they support, their peculiar function being that of "marking" the syntactic (or semantic) role of content words, rather than directly carrying semantic information. Moreover, they are the most frequent words in the language: only functional words can be found among the first 100 most frequent forms, in a frequency lexicon of Italian (ref 5). Consequently they are very efficient in providing information about both the local context and the global structure of the sentence.

The grammatical categories- The syntactic information provided by function words is usually negative, in that it excludes some possibilities while it is not sufficient to choose the only one fitting the context. The most profitable use of function words is possible when the ambiguity is specified: for example a word preceded by an article can be a noun, an adjective, an infinite verb, but if the specific ambiguity is noun vs. finite verb, which is often

the case for stress ambiguities, then the choice is deterministic; moreover, lexical peculiarities may prevent some sequences of categories and should be taken into account at least for very frequent ambiguities. A careful choice of the different syntactic roles and of the kinds of ambiguity classified in the system of grammatical categories is essential to the effectiveness of the contextual rules. The classification of function words should be very detailed in order to record all the information they can provide, including gender and number (allowing agreement tests) and the lexical peculiarities of their syntagmatic relations; e.g.: the preposition "in" should be kept in a separate class in order to recognize the locutions "in balia", "in seguito" and solve the ambiguities *bali'a vs. ba'lia, se'guito vs. segui'to*. For those function words which can play different roles in the sentence there must be contextual rules that solve the ambiguity where necessary. Demonstratives and indefinites must be assigned ambiguity symbols, because they can be adjectives or pronouns: the ambiguity should be solved in order to choose whether to de-accent them or not; they may be left ambiguous if the end is telling whether the preceding verb "to have" occurs as an auxiliary or not.

Definition of the grammar- The grammar, aiming at substituting to specific ambiguous descriptions of the syntactic role of a word the more defined description represented by a category, can be formally defined as a pair $\langle S, R \rangle$ where $S = A \cup C$ is the union of a set A of ambiguity symbols (including the symbol "*" standing for "any category") and a set C of category symbols (including symbols for punctuation marks), while R is a set of rules having the following form:

$$A_{c1} \dots c_n \text{ ---} \rightarrow C_i \text{ (} i \leq 1 \leq n \text{) / } S_1 \dots S_m \quad S_h \dots S_k$$

which means that the symbol representing the ambiguity: C_1 vs. ... vs. C_n can be substituted with the specific category C_i if it is preceded by the sequence of symbols $S_1 \dots S_m$ and followed by $S_h \dots S_k$ (each S_i being a category symbol or an ambiguity symbol, including "*"). The syntactic analysis of the input text is performed through the following steps. Each word is labeled with a category or ambiguity symbol, by dictionary look-up or morphological analysis. About 1,000 words are classified in the dictionary: 350 function words, 150 homographs, and exceptions to the phonological rules. The following are sample labelings:

(word in the dictionary)	"il"	----->	ARTICLE (category symbol)
(word in the dictionary)	"ancora"	----->	VERBNOUN (ambiguity symbol)
(word ending with -era')	"posera"	----->	VERB (category symbol)
(word ending with -ano)	"remano"	----->	VERBNOUN (ambiguity symbol)
("unknown" word)	"cielo"	----->	* (ambiguity symbol)

Next, for those words whose phonetic or prosodic realization depends on their syntactic role, the substitution of the ambiguity label with a category symbol is tried, via the application of the contextual rules R (ordered by "effectiveness"). The contextual analysis algorithm is simple: the rules are applied left-to-right to the labels corresponding to the words of the input text. Look-ahead is required in some cases, where the right context of the rule is a category symbol matching with a more general ambiguity symbol in the analyzed text.

The statistical ground of the rules- Once the system of categories, represented by the set S of category and ambiguity symbols, has been defined, the set R of contextual rules is determined by testing the relative distribution of categories in written Italian texts.

The text corpus amounts to about 380,000 words; it includes novels, encyclopedia items and the text corpus (C.E.E. documents; 100,000 words) analysed in the Esprit Project n. 860 "Linguistic Analysis of European Languages"; a segment of the corpus (150,000 words) has been grammatically analysed, each word resulting labeled with a category representing its syntactic role in the context (up to the detail of mode, time, number, gender); flexible statistical tools allow to test the behaviour of categories and ambiguities of the system S. Function words cover about the 47 % of the text corpus. The main goal is to determine the syntagmatic relations involving function categories. Given an ambiguity symbol $A_{c1}c2$ (a binary alternative is the typical ambiguity in our system) and a category or ambiguity symbol F representing a class of function words, the conditional probabilities $p_{c1}(F)$ and $p_{c2}(F)$ of the occurrence of C_1 and, respectively, of C_2 in the context of F are compared. If, say, $p_{c1}(F)$ (transitive and progressive) is significantly great and significantly greater than $p_{c2}(F)$, then the following rule will hold and will be included in R: $A_{c1}c2 \text{ ---} \rightarrow C_1 / F$

Table 1. Sample conditional (progressive) probabilities $p_c(F)$ The lesser is the ratio:

	NA	V	Rules
PRO	.000	.694	(1) $A_{na}v \text{ ---} \rightarrow V / PRO$
ESS	.232	.000	(2) $A_{na}v \text{ ---} \rightarrow NA / ESS$
ART	.871	.000	(3) $A_{na}v \text{ ---} \rightarrow NA / ART$
ART *	.323	.028	
	NA_{sm}	V_{3p}	
$ART_{sm} *$.291	.001	(4) $A_{na_{sm}}v_{3p} \text{ ---} \rightarrow NA_{sm} / ART_{sm}$

$p_{c2}(F)/p_{c1}(F)$, the more reliable is the rule; the greater is $pF(A_{c1}c2)$ the more effective it will be in solving the ambiguities of the class $A_{c1}c2$.

Table 1 reports as an example a set of statistical results relative to the ambiguity Noun or Adjective vs. Finite Verb, and the corresponding rules. It is appa-

NA=Noun/Adjective; V=Finite Verb; PRO=Proclitic; ART=Article; ESS=Verb "to be"; sm=singular masculine; 3p=third pers. plur.

rent that the high degree of detail and specialization of the category system allow the definition of more reliable rules: for example, while the ambiguity $ANA \vee$ can't be solved on the basis of the preceding occurrence of ART *, because of the frequent sequence ART NOUN VERB which is typical of the standard syntactic structure Subject+Predicate, the rule (4) which is adopted to solve the large class of ambiguity composed by those homographs which can be singular masculine nouns or adjectives and third person plural verbs is based on the ratio $p_{V3p}(ARTsm *) / p_{NAsm}(ARTsm *) = .003$ as the sequence ARTsm * V3p, that the lack of agreement prevents from being an example of Subject+Predicate, results to be far less plausible in Italian than ARTsm NA NAsm, which is a typical noun phrase.

APPLICATION OF THE GRAMMAR IN HOMOGRAPHY SOLUTION AND DEACCENTUATION

As examples of the actual use of the grammar, two specific applications, concerning homography solution and phonological word build up, are reported.

150 homographs are listed in the dictionary, including the 40 homographs appearing among the 6,000 most frequent words in Italian (ref 5). 12 ambiguity symbols are specifically designed to solve homographs ambiguities: to each about ten contextual rules are associated. The homographs which show the highest frequency of occurrence in the language ("subito", "seguito", "ancora", etc.) are represented by ad hoc ambiguity symbols, in order to take advantage of their lexical peculiarities (see above se'guito vs. segui'to). The application of the rules is ordered by effectiveness. If none of the rules applies to the context, the default choice criterion is the relative frequency of occurrence of the two alternatives in those linguistic contexts which are not represented by the rules. The rules have been tested on a corpus of 380,000 words. The correctly solved ambiguities are 1,062 on the 1,096 occurrences of homographs in the corpus: the error rate is 11% when the only frequency criterion is applied and 3% when contextual rules are added.

The function words which are de-accented (unless some contextual-structural conditions are verified) belong to the following categories: articles, prepositions, conjunctions, proclitics, subject pronouns, adverbs, possessive numeral indefinite demonstrative interrogative adjectives, auxiliary verbs, modal verbs, aspectual verbs. While the first four classes correspond to category symbols assigned by dictionary look-up, the others correspond to syntactic roles to be detected by contextual analysis: about 20 ambiguity symbols are involved, each associated to a set of about 12 rules. As function words may occur in sequence, look-ahead is often needed. When the ambiguity is not solved, the word is not de-accentuated, as a wrong de-accentuation may affect intelligibility. The de-accentuable categories cover a high percentage of the text; it turns out that more than one word every four is likely to be de-accentuated, the global effect on fluency resulting sensible.

The de-accentuation operative rules have the form: **if** (Ci & Cond) **then** dj
 where Ci is a category symbol and dj indicates which kind of de-accentuation must be performed on the word belonging to Ci when the condition Cond is verified. Typical contextual conditions affecting de-accentuation are: distance between stresses; length of the word; length of the following phrase; sentence modality. Lexical prosodic rules are defined for those words which show a peculiar prosodic behaviour, different from that of the grammatical class they belong to. Some words are affected by the presence of adjacent de-accentuable words; when functional adjectives occur in sequence some of them retain the accent, depending on lexical peculiarities. While, in natural speech, this fact is affected by the semantic context and on personal habits of the speaker, a generalization has been tried; rules have been defined based on a hierarchy stated on the set of adjectives depending on their tendency to lose the accent; the words "altro, stesso" turn out to be the most likely to retain the accent, while "questo, quello" the most likely to lose it.

CONCLUSIONS

The syntactic processor here described is embedded in a text-to-speech system for Italian implemented on Olivetti M24 Personal Computer (ref 6); it is composed of: a small dictionary including function words and exceptions to phonological rules; a list of word-endings and morphological rules; the grammar, that is a set of contextual rules of straightforward application, based on function words. This kernel can be developed in a more complex parsing system once a larger dictionary and a more time-consuming processing can be afforded.

The described approach could be extended to other languages: once a class of function words is singled out, contextual rules can be extracted by statistical analyses of written texts while language-specific 'prosodic rules' relating syntactic and phonological facts, can be stated by analyses of read texts.

REFERENCES

1. O'Shaughnessy, Proc. ICASSP, p. 1430 (1987)
2. Chomsky-Halle, The Sound Pattern of English (Harper & Row, NY, 1978)
3. Farnetani, Kori, Quad. Centro Studi Ricer. Fonetica Padova, p.287 (1983)
4. Bertinetto, Annali Scuola Norm. Sup. Pisa, Cl. Lett. Fil., serie 3, vol.15, p.581 (1985)
5. Bortolini & A., Lessico di frequenza della lingua italiana contemporanea (IBM Italia, 1971)
6. Vivalda, to be published on Olivetti R & T Review, n.7 (1987)