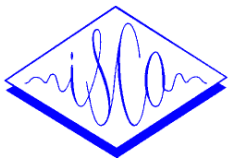

ETRW on Dialogue and Prosody
Veldhoven, The Netherlands
September 1-3, 1999

A framework to allow Dialogue Systems to generate Context-Sensitive Prosody

Pierre Larrey
IRIT - UMR CNRS 5505 - Paul Sabatier University
31062 Toulouse Cédex - France
Tél.: ++33 561 558 835 - Fax: ++33 561 556 258
e-mail: larrey@irit.fr

ABSTRACT

Recent progress of spoken dialogue systems now allow the extension of their abilities, and a richer interaction will need the use of language generation techniques in order to output appropriated messages. Additionally, prosody can indicate speech act, modality, relative salience of concepts, or play discourse functions. Thus, the same words can be pronounced differently according to context, and system's intentions. We introduce a framework linking different communicative goals to different prosodic patterns. We present a two-level of commands for French intonation : the first level (phonological), allowing a fine control of low-level prosodic properties of utterances, the second level (conceptual) made of dialogue act, information structure and enunciation schemes. We then expose our solution to the problem of mapping those two levels introducing the phonostrategies. A phonostrategy is a plan consisting of rules which map the discourse function of prosody and the prosodic commands, like macro-commands that are applied to sequences of symbols representing the concepts of the utterance.

1. INTRODUCTION

We are developing mixed-initiative dialogue systems but where the computer plays a prominent role : it prompts the user, who has the opportunity to utter non-requested information. But even in this kind of dialogues, misunderstandings are common and the system must often check the information provided by the user, handle its own comprehension failure and appropriately react to users' lack of cooperation [7]. Over the telephone, changes of textual content and/or prosody modification are the only resources the system can count on, and prosody is the major key to naturalness. We therefore decided to study the relationships of prosodic features in French and semantic or pragmatic information (which we will refer from now by "discourse" information) which is of the level of intention and is partly separated from the textual content. Our objective is to allow dialogue systems to generate appropriate prosody in several interaction contextes, and to vary the prosody of messages with the same text.

Prosodic features, like pitch, are capricious phonetic features : they expand over domains of various sizes, vary from speaker to speaker, from situation to situation, and have non deterministic acoustic correlations. A level of abstraction is thus required to capture similarities of prosodic patterns. This has been the objective of a wide range of studies which try to establish that "there is a phonology of intonation that is separate from the phonetics" [3]. On the other hand, studying and representing meaning and intention is a complex problem and the factors influencing the construction of the phonologic level are multiples and interplay : syntactic structure, lexical content, speaking style or rate, discourse function, thematic information, and many others (rhetoric relations, emotions, etc.). Determining what information is needed to compute prosody is still an on-going research subject. We believe theoretical linguistic requirements meet here the computational practical ones : a modular approach, involving successive autonomous levels of representation, with formal explicit rules mapping bi-directionally those different levels.

A phonetic level able to capture perceptually equivalent intonation has been implemented, based on syllabic categorization and symbolic coding, and above it, a set of phonological commands that can generate acceptable prosody with a reduced set of explicit implementation rules [5]. In this paper, we describe the further step and present a framework that is able to model some of the relations between conceptual representation of intention and the phonological level. Our proposal is to solve the complexity of this task, offering the possibility of writing prosodic macro-commands, whose implementation details are referred as phonostrategies. This work allows the implementation of a practical concept-to-speech system, in the simplified context of task-oriented dialogue. It integrates separate phonological and phonetic levels, where the rules of phonetic implementation as well as tune-text association can be expressed in a declarative way, and thus being re-usable and modifiable.

2. DIALOGUE RECORDINGS

We recorded and studied human-human dialogues under particular conditions. We decided to study one single speaker phonologic strategies and then we made a professional receptionist answer to queries from users (people from the laboratory) over the telephone. Queries were made about train schedules in the area of Toulouse. Observers could produce a specific noise to prevent the caller or the receptionist from hearing what was uttered, to simulate speech recognition errors or speech synthesis lack of intelligibility. The caller had to plan a trip between two cities, a day and an approximate time of travel. The receptionist received several directives : he should performs checks of what the caller said, and always ask questions to keep the initiative in the dialogue. He was also

given with particular strategies to handle misunderstandings or dialogue failures : try to repeat the same words in case of incomprehension, only trying to be more intelligible ; in case of persisting misunderstanding he should propose an explicit list of choices or utter a yes/no question. 15 dialogues (317 utterances) were recorded (training dialogues were not recorded) over several days.

This corpus (166 utterances) is short and does not reflect *real* situations, but we think however the communicative functions of prosody are still realized, and we must mention that we do not aim to characterize with exactitude the discourse functions, we rather propose a framework in which the prosodic properties of communicative goals can be expressed and later investigated and assessed with more representative corpora, with different speakers, speaking styles and situations. Thus, for this goal, this corpus has been judged appropriate.

3. PROSODIC LABELING

The 166 utterances have been labeled with the prosodic system developed in [5], that we briefly summarize now.

The phonetic TMD level (Tone Melody Duration) has been validated by perceptual tests with listeners in a procedure of analysis/resynthesis [5]. It consists of syllabic categorization over three axis : the total syllabic duration (7 categories), the pitch level (3 categories) and pitch pattern (tone : 8 categories, combinations of rising and falling tones). Reset of numeric parameters representing top line and register mean can occur after any silence. Provided with the segmental content this level can be fully automatically extracted from speech. The phonological level PSML (Prosody Synthesis Markup Language) is an implementation of Jun an Fougeron Model [4] : the lowest prosodic domain is the Accentual Phrase, which has 4 underlying tones (LHiLH*) ; this domain can be embedded in intermediate phrases (with final boundary tone L or H), which in turn is embedded in Intonational Phrases. The container tags do not need to be closed, as they are part of a hierarchy. Tones are indicated by the empty element <tone> that specifies the phonological tone on the next syllable. An IP is coded with an <ip> tag, an ip with an <inter> tag. Within one ip, a particular AP could receive a focus attribute that could enhance the phonetic level of one or both Hi and H* and de-accentuate the following APs. A special tag for pauses is also part of the model as well as a categorized representation for register (Top, Medium, Bottom, relative to particular values that can be adapted to a speaker) and range (Wide, Normal, Narrow). These categories applies for each ip, and when indicated on one IP, are set for each contained ip.

A manual labeling into PSML tags has been done by one expert using a visual representation of an F0 curve and listening to the original speech. A linguistic representation of the prosody of the 166 utterances have then been obtained. This representation allows us to generate speech where the prosodic similarity to the original is perceptually acceptable in most of the cases [5]. The problem of mapping intention and prosody is then reduced to the mapping with the features of this representation.

4. DISCOURSE LABELING

To characterize the prosody of intentions and propose computational means for the mapping, a multi-tier labeling system has been decided to represent intentions. The first tier consists in dialogue acts which summarize the intention of a stretch of speech. But the structure of the dialogue acts is generally varied and the intonation pattern highly depends on this structure. The second tier of labeling labels segments of speech and reflects the information structure by identifying particular typical constituents of spontaneous French (by now referred as discourse segments no to be mistaken with stretches of speech longer than utterances). Finally, a syntactic-semantic labeling has been done using conceptual segments that are semantic units showing a syntactic cohesion.

4.1. Dialogue moves labeling

The dialogue acts are described with the coding scheme for conversational games (exchanges) used to describe the English HCRC Map Task Corpus [1] and are referred as moves. The conversational moves considered are initiation ones : INSTRUCT, EXPLAIN, ALIGN, CHECK, QUERY_YN, QUERY_W, and response ones : ACKNOWLEDGE, CLARIFY, REPLY_Y, REPLY_N, REPLY_W ; and preparation one : READY. Although this coding scheme could cover the acts occurring in our corpus, we found it useful to refine some of the moves, and add others which have a more precise purpose and a specific prosodic pattern. These moves are referred as NO_ACK, NO_REPLY, NO_READY, LISTEN, THANKS, PLEASE, GREETINGS, CLOSURE, QUERY_ALT, REPLY_ALT. Of course further refinements could have been done but it would have resulted in a loss of generalization.

Sometimes the speaker did not hear what the other said and he then utters a NO_ACK move (e.g. “Excuse me”) which functions to inform the hearer of the failure. At some point, the speaker does not have the answer to a specific question and replies a NO_REPLY move that contrasts prosodically with ordinary REPLY moves. One example is “There is no train at this time”. Opposed to a READY move that leaves the opportunity to the hearer to initiate a new game (exchange), the NO_READY move beg the hearer to wait for the next moves. Typical examples of this moves are “Please wait”, “Hold on” and so on. Our speaker as well often uttered what we called a LISTEN move which purpose is to draw hearer’s attention to his next moves, for example : “Well”, “So”, “Let us summarize”. To these moves, we found useful to add typical polite business telephony conversation acts : THANKS (“Thanks for calling”), PLEASE (“Please”), GREETINGS (“Welcome on this train schedules server“), and CLOSURE (“Goodbye”). All these moves have their own prosodic characteristics [2]. We decided also to refine the QUERY_W and REPLY_W moves to handle the special characteristics of alternatives : QUERY_ALT (“Would you like to travel in the morning or in the afternoon ?”), and REPLY_ALT (“departure Toulouse-Matabiau 16h07 arrival StSulpice 16h38 or departure Toulouse-Matabiau 17h00 arrival StSulpice 17h20”) that we found more relevant to code as one REPLY_ALT move than

two REPLY_W moves. We decided not to allow overlapping acts.

4.2. Conceptual segments labeling

A conceptual segment is a stretch of speech referencing a concept, i.e. it is a semantic unit. In task-oriented dialogues, it is generally possible to segment any utterance into such segments. These units have in French a syntactic and morphologic cohesion that makes their concatenation easy for a language generation purpose (with simple feature-matching rules).

1. alors départ Boussens 10h26 arrivée 11h puis départ de Toulouse à 11h01 … ah non ça risque d'être juste celui là . bon notez-le quand même départ Toulouse 11h01 arrivée Montauban à 11h28

   (so departure Boussens 10h26 arrival 11h then departure Toulouse at 11h01... oh no it will be too late. well you can note it anyway departure Toulouse 11.01 arrival Montauban at 11.28)

   LINK DEPARTURE PLACE TIME ARRIVAL TIME THEN DEPARTURE PLACE TIME NO COMMENT LINK NOTE ANYWAY DEPARTURE PLACE TIME ARRIVAL PLACE TIME

Moreover, French is well known for having a final primary accent that segments speech into semantic groups. The correspondence between H* delimited groups, namely Accentual Phrases, and conceptual segments is obvious. Our corpus confirms this hypothesis : a conceptual segment is composed of one AP, sometimes several, and AP merging (two segments consisting in one AP) is a particular operation that occurs under particular conditions. It thus permits separating prosody generation task (from intention to phonological representation) into global property determinations (above the conceptual segments : discourse segments and dialogue moves influencing phrasing, register selection, declination, etc.) and local properties determination (specific to conceptual segments : tone association, initial accent realization, AP merging, etc.). However, the segmentation into semantic units is rather non deterministic (two units can sometimes be considered as one), to ensure the maximum consistency to our labeling we decided to respect syntactic boundaries and generally start a conceptual segment with a function word. When necessary, a new concept (i.e. a new conceptual segments class) was created, even if not matching a system response generation purpose (see COMMENT concept in the example 1 above). Conceptual segments are the like branches of our integrated discourse representation of intention, but it seems rather evident that a flat sequence of concept does not give the information on how they are prosodically uttered, this is the purpose of labeling discourse segments that constructs a hierarchy above the concepts.

4.3. Discourse segments labeling

A recent study of spontaneous French [6] proposed a new approach to the segmentation of speech into functional discursive units. This approach is based on the co-enunciation concept : the speakers always considers the hearer as a potential information giver and make assumptions about which information or point of view can be shared or not with him (notion of convergence or divergence). This concept is obviously fundamental for dialogue where negotiations continuously occur. A simple structure for utterances consists in a preamble, a rheme and an optional postrheme. This sequence forms a basic oral paragraph, which is the domain for declination in French. Several sequences of preambles and rhemes can follow each other to form a complex paragraph. Also any rheme preceding another rheme becomes its preamble. The function of the preamble is to construct a referent upon which the speaker supposes a high degree of convergence, then the rheme segments expresses what the speaker considers as its own personal contribution to the co-enunciation. The function of the postrheme is a rupture of the co-enunciation process, which occur when a speaker utters something upon which the hearer can not have a point of view. The preamble is composed of several constituents that progressively focus on a particular discourse content : a link ("then", "well") , a point of view expression ("in my opinion", "for me"), a modus ("maybe", "surely"), several frames which progressively reduces the semantic fields and a lexical disjoint support that builds the topos that will be referred to in the rheme. It corresponds to a progressive construction of a discourse content characteristic of oral conversation. Generally one segment in the preamble assumes several functions. An example of an oral paragraph is :

2. enfin, d'après moi, pour Grisolles Gaillac le samedi, le plus simple, c'est le car (litt. : well according me for Grisolles Gaillac on Saturdays the most simple it is the bus)

This oral paragraph is almost complete with a long preamble consisting of a link ("enfin"), a point of view ("d'après moi"), a two pieces frame ("pour Grisolles Gaillac", "le samedi"), a lexical disjoint support that also plays the role of modus ("le plus simple") and finally a short rheme ("c'est le car").

Our corpus showed significant variants of this canonical scheme because its specificities are : purely informative exchanges due to task oriented dialogue, multiples query-reply : repartition of paragraphs between interlocutors (the co-enunciation concept plays here its full role), elliptic constructions, lack of personal implication (formal dialogue), implicit information, structure closer to written text (mental reading), reading of some pieces of information (schedules and stations). So, a big majority of paragraphs have short preambles, often limited to a link and a frame, rheme-only utterances where frequent (elliptic), and structure of utterances was sometimes closer to written French. We had to handle rheme decomposition between interrogative modus (im), verbal group and oral punctuation mark (ponct) and also allow im to occur in preambles and to permit rheme-preamble inversion,

see examples. (generally the preamble tag is obvious and can be omitted). Examples below show some of these specificities.

3. <im> désirez-vous <im> <rheme> un autre renseignement ? </rheme>

4. <para> <frame> <im> avez-vous dit que vous souhaitiez </im> </frame> <rheme> <gv> partir </gv> <im> vers midi ? </im> </rheme> </para>

This discourse segment coding scheme has been chosen because the segments have specific prosodic characteristics that we will explore in detail in the next sections and that allows to determine major part of the prosodic structure, being a finer model than the usual theme/rheme decomposition.

4.4. Special Enunciation Schemes

A complete description of the information needed to compute prosody is evidently out of reach at the present time : rhythmic considerations, styles, rhetoric relations, speaker specificities, emotions or attitudes are ones of the numerous factors still to be included in a prosody generation process. However, it has been observed that our speaker made use of conversational stereotypes in his enunciation process that seemed orthogonal (although interacting) to dialogue moves and informational discourse structure. These stereotypes have been referred as special enunciation schemes and we decided of a set of tags to label our corpus in order to handle those schemes and find their prosodic correlates, to generate them. Examples of such special enunciation schemes found in our corpus include : listing, emphasis, alternative, menu, right dislocation, left dislocation, incise, hyperarticulation. It has been observed that those stereotypes act at different levels : some of them modify the information structure, while others modify directly the phonetic level, but generally they have phonological correlates.

To summarize, we can say that dialogue acts represent *purpose*, discourse segments, *structure*, conceptual segments, *content* and enunciation schemes, *manner*.

5. PHONOSTRATEGIES

As study of the relationships of discourse and prosodic labels is presented here. We show how for each dialogue move, the information structure influences the intonation pattern. These mappings can be formally written as rules representing phonological strategies (phonostrategies) that a speaker can choose to achieve a communication goal. Some examples of phonostrategies are given as well as the mechanism that generates the PSML representation from the discourse labels.

5.1. Prosody properties of discourse labels

Characteristics of discourse segments. Observing the prosodic and conceptual levels of representation some systematic properties of discourse segments appeared. For example a preamble usually forms an intermediate phrase with a H- boundary tone (rule 1), also when grouped with a rheme before another rheme (rule 2). Generally a rheme forms an intermediate phrase in final position (rule 3). This single example illustrates these three observations :

5. alors vous avez un train à 11h arrivée à Toulouse à 11h 54
   link H- rheme H- frame H- rh L-L%
   (so you have a train at 11 arrival Toulouse at 11.54)

One can argue that a simpler rule saying “non-IP terminal ip receive a H-“, but in our corpus we found some L- boundary tone occuring at IP-internal position , as in :

6. vous désirez voyager en semaine ou le week-end ?
   L Hi L H* Hi L H* Hi L H* L H*
   L- H- L-L%

   (you’d like to travel during the week or in the week-end ?)

The L- tone here makes no doubt because the H* tone is realized phonetically by a low-pitch melodic syllable category.

Another clear example of systematic relationship between the prosodic labels and the discourse segment type is the postrheme, which is uttered at the bottom register of the speaker and has a flat intonation, that can be interpreted as a narrowed range (rule 4).

Intonation of Dialogue Moves. In our corpus, we found that the intonation of dialogue moves partly depends on the informational structure of them. For instance, in a query if the interrogative modus precedes the rheme of vice-versa, the melodic pattern will be different. In the corpus, we found some moves that show the same structure and others that show differences in regard to the information structure. We present her some examples.

The LISTEN move was found sometimes forming a IP on its own (see 11) and sometimes being the first ip of an IP. In the first case the IP boundary is always H-L% while in the second case the ip ends by a H- and tone of the IP depends on the next moves, indeed this last case occurs when the move is realized by a link discourse segment (see 10).

7. *alors* départ Lavaur à 12h39
   H- H- L-L%
   (so departure Lavaur at 12.39)

8. *récapitulons* : vous voulez … etc.
   H-L%
   (let us summarize. you want … etc.)

All GREETING moves show the same characteristics : H-L% boundary tone, a wide range, and consisting of only one IP. All EXPLAIN moves in the corpus have L-L% boundary tone, and a narrow range. We characterized this way all the dialogue moves of our corpus, more examples will follow in the next section, written as phonostrategies.

Enunciation schemes. The labeling of the intonation schemes allowed us to express them using PSML commands. One example is the LISTING schemes that associates one intermediate phrase with a H- boundary tone with each item.

Also an emphasis tag, was sometimes realized with a pause, and a focused AP or an ip restructuring with a focused AP. Hyperarticulation produced pauses between syllables, shorter APs and systematic Hi. Right dislocation appeared in our corpus to share the same properties than the postrheme, while left dislocation acted like a preamble (a lexical disjoint support). Incises, not surprisingly, were uttered by our speaker as intermediate phrase having a low register, with narrowed range, but with a final H- tone.

5.2. Mapping rules

Intonational and intermediate Phrases. Using the markup schemes for labeling prosody and intention, we are able to express the observations of our corpus using phonostrategies. A phonostrategy is a rule involving dialogue moves tags or discourse tags, or enunciation tags (section 4.4.) containing PSML tags, it can also be viewed as macro-command of PSML commands. For example, recalling the rules numbered 1 to 3 above, we can write them as phonostrategies :

<pre><inter final=H-> S </inter> </pre> (rule 1)
<inter final=H-> <pre> <rheme> </inter> [<rheme>] (rule 2)
<rheme><inter> S </inter> (rule 3)

Before explaining the syntax of the phonostrategies and the mechanism that combines them let us present the phonostrategies representing the CHECK moves that we found in our corpus.

(a) <CHECK><ip final=L-L%> <link> S <./ip> </CHECK>
(b) <CHECK><ip final=H-H% reg=top> <rheme></ip> </CHECK>
(c) <CHECK><ip final=L-L%> <inter final=H- reg=top> <im> </inter> S </ip> </CHECK>

Examples illustrating those mappings are :

declarative preceded by a link (a)
9. donc avant 8h du matin (so before 8.am)
L-L%

single rheme (b)
10. aujourd'hui ? (today?)
H-H%

interrogative modus at the start of the move (c)
11. est-ce que vous partirez de Toulouse ?
H- L-L%
(will you leave from Toulouse?)

We give other examples of phonostrategies obtained from corpus labeling :

<ACK> <ip final=L-L%> S </ip> </ACK>
<EXPLAIN><ip range=narrow final=L-L%> S </ip> </EXPLAIN>
<ALIGN> <ip final=L-L%> S </ip> </ALIGN>
<ALIGN> <ip final=L-H%> S <im> </ip> </ALIGN>
<ALIGN> <ip final=L-L%> S <inter final=H-> <im> </inter> S </ALIGN>
<QUERY_YN> <CHECK> <QUERY_YN>
<QUERY_YN> <ip final=L%> <inter final=H-> <rheme> </inter> <inter final=L-> <im> </inter> </ip> <QUERY_YN>
<QUERY_YN> <ip final=H-H% reg=top> S </ip> </QUERY_YN>
<NO_READY> <ip reg=top final=H-L%> S </ip> </NO_READY>
<LISTEN> <inter final=H-> <link> </inter> S </LISTEN>
<REPLY_W> <ip final=L-L%> S </ip> </REPLY_W>

The mechanism that generates a PSML string using these rules is two stage : based on a string-matching unification algorithm for the first stage (above the level of the conceptual segments, i.e above the AP), then operators modify AP structure (AP merging, initial accent specification, focus. The symbol S represent any sequence of discursive labels (not an act). A discourse tag which is not closed (e.g a <rheme> without </rheme>) is unified with a complete segment of the same type in the input (see the preamble and the rhemes in rule 2). The PSML tags are not considered for the unification but remains in the result, tags between [] are used for the unification with a rule but remains available for the next unification. A hierarchy of the tags is used for the unification process : utt > para > acts > discourse > PSML. Default rules are applied : default boundary IP tones is L%, compound IP boundary tone T-T% is prevalent over last ip tone, any free domain at the left of a higher domain is included in it. The system search for a phonostrategie that match the input act, using context (contained discourse segments), unifies S, then goes on until all the discourse marks and the dialogue act tags disappeared.

Accentual Phrases. The conceptual segments, stored as AP whose tones appear as in citation form, are finally included when the phonostrategies have applied. Phonostrategies including AP-level rules are implemented using operators. For example : SCHWA operator restructures the AP so that the optional mute e is realized, the SYLLABIFY operator inserts a pause between each syllable, the NO_SEC suppress the Hi tone, the FOCUS operator trigger a focus=yes attribute for the AP. In the rules of this level the symbol C represent any sequence of concepts and c unifies only with one concept, the operators of course are not unified but inserted. Examples of rules are given with the enunciation schemes.

Enunciation Schemes Coding. The enunciation schemes involve rules at all the levels : AP, ip and IP. Here are some examples :

<LIST> <item> <inter final=H-> S </ip> </LIST>
<LIST> <IP final=L%> <item> <inter final=H-> S </inter> <LIST> </LIST>
<HYPER> syllabify,schwa : c <HYPER> C </HYPER> </HYPER>
<EMPH> <HYPER> </EMPH>
<EMPH> focus : c </EMPH>
<EMPH> <pause> focus : c </EMPH>
<INCISE> <inter reg=bottom, range=narrow final=H-> S </inter> </INCISE>

We show now a complete process of generation of a PSML string from a conceptual input with act and discourse information.

INPUT : <para > <LISTEN> <link> LINK </link> </LISTEN> <REPLY_W> <frame> DEPARTURE </frame> <rheme> TIME </rheme> <frame> ARRIVAL </frame> <rheme> TIME </rheme> <rheme> PLACE </rheme> </REPLY_W></para> <para><act type=QUERY_YN> <im> Q_OK </im> </para>

1-unification act after act : with <LISTEN> <inter final=H-> <link> </inter> S </LISTEN>, then with <REPLY_W> <ip final=L-L%> S </ip> </REPLY_W> and with <QUERY_YN> <ip final=H-H% reg=top> S </ip> </QUERY_YN>

RESULT : <para> <inter final=H-> LIG </inter> <ip final=L-L%> <frame> DEPARTURE </frame> <rheme> TIME </rheme> <frame> ARRIVAL </frame> <rheme> TIME </rheme> <rheme> PLACE </rheme> </para> <para><ip final=H-H% reg=top> <im> Q_OK </im> </para>

2-unification of discourse segments (the <pre> tag includes <frame>) with rules 1 and 2

RESULT : <para> <inter final=H-> LIG <ip final=L-L%> <inter final=H-> DEPARTURE TIME <inter final=H-> ARRIVAL TIME <inter final=H-> PLACE </inter> </para> <para> <ip final=H%> <inter final=H-> Q_OK </para>

3- insertion of utt tags, removing empty discourse tags (para), insertion of APs from the lexicon in place of concepts, shifting of the initial IP to embed the free ip at its left

RESULT : <utt> <ip final=L-L%><inter final=H-> <ap> <tone pa=L> a <tone pa=H*> loR <inter final=H-> <ap> <tone pa=L> de <tone pa=H*> par <ap> <tone pa=Hi> Hi <tone pa=L> t9R <tone pa=H*> ve~<inter final=H-> <ap> <tone pa=L> a Ri <tone pa=H*> ve <ap> <tone pa=Hi> n9 v9R <tone pa=L> ve~ (t@) <tone pa=H*> tRwa<ap> <tone pa=L> a <tone pa=Hi> al <tone pa=H*> bi<ip final=H%> <inter final=H-> <ap> <tone pa=L> se <tone pa=H*> bo~ </utt>

alors	départ	8h20	arrivée	9h 23	à	Albi,						c'est bon ?	
L H*	L	H*	Hi	LH*	HiLH*	Hi	LH*	L	HiH*			L	H*
	H-			H-			H-		L-L%				H-H%

(so departure at 8.20 arrival 9.23 in Albi, ok?)

6. RESULTS AND CONCLUSION

The 166 utterances of the corpus have been separated in two parts in such a way that dialogue moves repartition is identical in both corpora. The first part, corpus A, has been used for the study presented in the preceding section. The phonostrategies and mapping rules have been applied on both. Results on corpus A show the validity and the consistency of the algorithm and of the framework, while results on corpus B measure the reliability of the rules we have designed. Error rates (insertion, deletion, wrong tone) showed for corpus A nearly no errors for IP and ip phrasing (2.5%), some errors for boundary tone prediction (6,3%) and more errors at the AP level (insertion of H* and of Hi, wrong position of L tones, about 14% on average), this was because the citation form of the conceptual segment was always used (this however increase the intelligibility and decrease the speaking rate). On corpus B IP and ip-levels error had an average of 9% and AP-level error stayed at 15,3%. This results are encouraging and we are now investigating the phenomenon of deaccenting and AP merging.

We have characterized, using explicit and formal rules, the prosodic properties of discourse segments using right and left context. The properties of the dialogue acts have also been characterized, depending on the structure of the discourse segments they are made of (informational structure). One advantages of the phonostrategies is that they can represent various levels of details of what is known about the segments (from a simple IP-final boundary tone, to a decomposition in ip, a specification of register or range). A generation process of prosodic commands using the phonostrategies has been described and produces consistent results. This can be due to what we have observed in our corpus : IP phrasing, register, range and boundary tones seems to be mainly influenced by the dialogue moves ; the phrasing of intermediate phrases, their boundary tones, and sometimes their register and range depends usually of the informational structure of the act : preamble vs. theme repartition ; AP phrasing verified our hypothesis (although AP merging occurs), while the presence of an initial accent Hi and of the position of L tones seems to obey to constraints that we have not modeled (AP merging). Or maybe the gap between phonetic realization and underlying phonological structure is responsible of the distorsion of what was labeled and what was generated. The speech generated is of good quality and is similar to the original recordings, so our framework proved to be usable in a spoken dialogue systems. Moreover, the set of rules remains open and modifiable to revisions or precisions that would result from larger and more representative corpora.

7. REFERENCES

1. Carletta, J., Isard, A., Isard, S., Kowtko, J., Doherty-Sneddon, G., and Anderson A. «The coding of dialogue structure in a corpus», *Twente Workshop on Language Technology on Corpus-Based Approaches to Dialogue Modelling*, Twente, The Netherlands, 1995.

2. Douglas-Cowie, E., and Cowie, R. «Macrostructures in Prosody : the Case of Phonecalls», *ESCA Workshop on Intonation*, Athens, Greece, pp 103-106, 1997.

3. Inkelas, S., and Leben, W.R. «Where Phonology and Phonetics Intersect : the case of Hausa Intonation»,in : *Papers in Laboratory Phonology : Between the Grammar and the Physics of Speech*, Cambridge, Cambridge University Press, UK, pp 17-34, 1990.

4. Jun, S.A., and Fougeron, C. «The Accentual Phrase and the Prosodic Structure of French», *Proceedings of the XIIIth ICPhS*, Stockholm, Sweden, pp 722-724, 1995.

5. Larrey, P., Vigouroux, N., and Perennou, G. «Synthesizing French Prosody form a Phonological Representation», *XIVth ICPhS*, San Francisco, USA, 1999.

6. Morel, M.A., and Danon-Boileau, L., *Grammaire de l'intonation*, Bibliothèque de Faits de Langues, OPHRYS, Paris, France, 1998.

7. Tatham, M.A.A., and Morton, K. «Speech Synthesis for Dialogue Systems», *ESCA Workshop on Spoken Dialogue System*, Visgø, Denmark, pp 221-224, 1995