Studies of Emotional Expressions in Oral Dialogues: towards an Extension of Universal Networking Language

Mutsuko Tomokiyo,[1] Gérard Chollet,[2] Solange Hollard [3]

[1] GETA-CLIPS-IMAG & ENST
BP 53 38041 Grenoble cedex 9 France
mutsuko.tomokiyo@imag.fr ; tomokiyo@tsi.enst.fr
[2] ENST
46 rue Barrault, 75634, Paris
chollet@tsi.enst.fr
[3] GEOD-CLIPS-IMAG
BP 53 38041 Grenoble cedex 9 France
Solange.Hollard@imag.fr

Abstract. Emotions entail distinctive ways of perceiving and assessing situations, processing information, and prioritizing and modulating actions [24]. The paper aims to study theoretical and pragmatic aspects of emotions and to propose a semantic representation of emotions for oral dialogues, based on an analysis of real-life conversations, telephone messages and recorded TV programmes, focusing on a relationship between prosody and lexeme for the purposes of a speech to speech machine translation. The semantic representation is made, by using the **U**niversal **N**etworking **L**anguage (UNL) formalism, in a way where lexeme, phatics, gestures, prosody and voice tone are taken into account at the same time.

1 Introduction

This work has been carried out in a continuation of "VoiceUNL" [21], which is one of subprojects of the "LingTour" [1] project. "VoiceUNL" is an extension of **U**niversal **N**etworking **L**anguage (UNL), which is a text-oriented formalism of semantic graphs,, to oral dialogues.

As for **s**peech to **s**peech **m**achine **t**ranslations (SSMT) or man-machine interactive systems, the detection and generation of emotions are an important issue from the viewpoint of the naturalness of dialogues [7], because *emotion entails distinctive ways of perceiving and assessing situations, processing information, and prioritizing and modulating actions* [24]. It's the key reason for proposing a semantic representation of emotions.

In this paper, section 2 is devoted to previous emotion studies mainly focussed on prosody: a survey of existing approaches to emotion detection and generation, theo-

[1] The Lingtour project was launched in 2002 by the partnership which consists of TsingHua University (China), Paris 8 University (France), INT (France), ENST-Paris and Bretagne (France), and CLIPS (France). One of the objectives of the projects resides in R & D to enable multilingual-multimedia MT on user-friendly tools [1].

Universal Network Language: Advances in Theory and Applications.
Research on Computing Science 12, 2005, pp. 286–299.

retical approaches to emotions, emotion definition, and emotion categories. In section 3, we investigate our corpus, detect emotion categories and emotion eliciting factors, and extract emotional expressions from it. In section 4, after having introduced UNL briefly, we propose a semantic representation of emotions within the UNL formalism, by adding tags representing speech dialogue properties to UNL to suit it to SSMT.

2 Previous Studies into Emotions

2.1 Existing Approaches to Recognition and Generation of Emotions

Much work has been carried out in detection and identification of emotions in written texts or oral dialogues for various applications. Existing approaches are grouped into three types:

- observation of paralinguistic elements such as prosody, facial and body movements in spoken languages,
- detecting lexical items expressing emotions by using a shallow word match parser or by physiologic manual evaluation and defining emotions by the distance between two emotions according to the distance values given [7][44], or
- discovering syntactic and lexical patterns in the text that allow emotion tagging [6].

Our approach employs a method where spoken language properties such as lexeme, gestures, prosody, etc. are recognized, translated and generated, since one objective of our emotion representation is SSMT using the UNL framework.
In fact, in order to determine the type of emotion, these elements are taken into account at the same time, because the same variable can express different classes of emotions. For example, an increase of elocution speed or the rising tone can indicate happiness as well as anger [22] [36].

2.2 Theoretical Approaches to Emotions

Cornelius mentioned there would be *four of the most influential theoretical perspectives and research traditions in the study of emotion in the past 125 years or so* [28] without citing Greek philosophers like Plato and Aristotle, nor French ones like Descartes. These four perspectives are the Darwinian , Jamesian [30], Cognitive , and Social constructive perspectives.

In the Darwinian perspective, emotions are considered fundamental because they represent survival-related patterns of responses to events in the world that have been selected in the course of our evolutionary history.

James [30] says that *the bodily changes follow directly the PERCEPTION of the exciting fact, and that our feeling of the same changes as they occur IS the emotion.*
In the experimentations of Levenson and al. [29], subjects received muscle-by-muscle instructions and coaching to produce facial configurations for anger, disgust, fear,

happiness, sadness, and surprise while heart rate, skin conductance, finger temperature, and somatic activity were monitored. Results indicated that *voluntary facial activity produced significant levels of subjective experience of the associated emotion, and that autonomic distinctions among emotions were found both between negative and positive emotions and among negative emotions.*

Consequently, the Jamesian group considers emotions as the expressive process of affective programmes which activate different subsystems.

Cornelieu pointed out that there is a considerable crossover between the Darwinian and Jamesian traditions in psychology. Its main point is that the bodily responses are associated with emotions. Thus, his suggestion [28] being based on the Levenson's experiments enables us to assume that the prosody can be also considered as one of these activated systems by one of affective programmes.

In the Cognitive perspective, as mentioned by Arnold [31], *thought and emotion are not separable, and all emotions are seen… as being dependent on the process by which events in the environment are judged as good or bad for us* [28]. Hence, every emotion is associated with a specific and different pattern of judgments of the worth, value, or condition of something, called appraisal.

The Social constructivists like Haviland [32] or Averill [33] assume emotions are cultural products that owe their meaning and coherence to learned social rules. For instance, anger plays a positive and constructive role in our social relationship, because, on the one hand, anger is generated only when one is intentionally wronged, and on the other hand, anger depends upon culture one belongs to.

Randall [8] states that most cultures have emotions and emotional vocabularies that have two components: a universal element, and a component or parameter that is peculiar to the beliefs and values of that culture. Hence, for the social constructivists, culture plays central role in the organization of emotions.

We could regard emotions a processes, consisting of several components from these perspectives: physiological, cognitive, sociomotivational and 'action tendency' as Scherer mentioned [40].

2.3 Definition and Emotion Category

Randall [8] defines emotion as a feeling that has been caused by certain beliefs, directed toward a primarily conceptual and non-perceptual target that typically produces some physiological, behavioural, or cognitive effect.
Cowie [27] mentions there are two different senses for the word 'emotion':

- The first sense uses the word, in plural form, *to refer to entities – natural units that have distinct boundaries, and that can be counted*.
- The second sense uses the word to *refer to an attribute of certain states. That is the sense that is involved when we say that somebody's voice is tinged with emotion.*

Our main concern to 'emotion' is one in the first sense, so our primary task is to obtain a list of emotion categories.

How many and what kind of emotional states are expressed in general dialogues?

Ekman [11] mentioned that there would be a linking of a second emotion with an initial emotion, and emotions rarely occur simply or in pure form. There is, however, quite general agreement on the so-called 'big six' as the initial emotion: fear, anger, happiness, sadness, surprise and disgust [28].

In OCC [9], there is an assumption that there are three major aspects of the world, namely, events, agents, or objects. Emotions *are valenced reactions, and any particular valenced reaction is always a reaction to one of these perspectives on the world*. Emotion types in OCC include 'happy' for resentment, gloating 'pity', 'hope', 'fear', 'joy', 'distress', 'pride', 'shame', 'admiration', 'reproach', 'love', 'hate', 'gratification, 'remorse', 'gratitude', and 'anger'.

Plutchik [10] believes that emotions are like colours. Every colour of the spectrum can be produced by mixing the primary colours. His "emotion's wheel" consists of eight primary emotions: fear, surprise, sadness, disgust, anger, anticipation, joy, and acceptance, and he lists 142 categories as second order emotion.

The emotion study group in Southern Kings Consolidated school [20] classifies the emotions as follows: Thankfulness, Envy, Disgust, Worry, Kindheartedness, Stress, Boredom, Sadness, Loneliness, Bravery, Paranoia, Optimism, Stubbornness, Fear, Anxiety.

Morita [41] classified Japanese emotional words into 40 categories, and it contains negative or positive judgment, sense to color or sound, psychological reactions, etc.

We extract and classify emotions by investigating our corpus.

3 Dialogue Corpus Analysais

3.1 Corpus

After surveying the theoretical aspect of emotions from the literature and available research papers on emotions and speech, we have developed in our first approach to emotions a corpus, which contains:

a. 30 minutes of English instruction programmes on TV,
b. a 40-minute French TV interview [5],
c. 5 hours of real-life vocal messages left on a telephone answering machine, sent from medical stuff to a group of computer engineers in a French public hospital [13],
d. 1 hour of real-life telephone conversations between administration staff of a French university [12] and
e. 6 basic conversations on transport in English, French, Japanese and Chinese [25].

We mainly used a. c. d. in the corpus.

3.2 Emotion Categories

We analyzed the corpus, either to fix emotion categories, or to find emotion eliciting factors.
We previously prepared a working sheet[2] to transcribe recorded material while listening to them or watching them, and checked the lexemes concerning emotional feeling, emotional prosody, emotional gestures, etc.

For emotion categories, we found the following categories in our corpus including the big-six we mentioned above: happiness, sadness/disappointment, disgust, surprise, fear, anger, irritation, hesitation/uncertainty, anxiety, and neutral.
These category names are used later to describe emotion states of the speaker as well as to annotate lexemes having emotional content in a semantic representation of emotions.

3.3 Emotion Eliciting Factors

For emotion eliciting factors, the followings are the major ones in our corpus :

- lexemes (sad, happy, etc.)
- phatics (ah, hein, etc.)
- prosodic cues (fast, slow, strong, etc.)
- voice (noisy, soft, etc.)
- gestures (movement of hands, mouth, eyes, etc.)

As an example, “No!” in the example

Victor - “May I smoke?”
Victor’s father - “No! you may not, Victor” [5]

expresses Victor’s father’s surprise, because Victor is a small boy.
So, the surprise can be represented by the lexeme “No!”. However, at the same time, on TV, the father also made a grimace while saying “No!”. Thus, surprise can also be expressed by the movement of the eyebrows and the voice tone.
From the example, we can suppose potential emotion expressions: ‘happiness’ is expressed by lexemes, phatics, prosody, voice, hand movements, mouth and/or eyes ; ‘sadness/disappointment’ is done by lexicon, phatics, prosody, voice, mouth and/or eyes ; ‘disgust’ is expressed by lexemes, phatics, prosody, voice, mouth, eyes, eyebrows and/or shoulder movements ; ‘surprise’ is expressed by lexemes, phatics, prosody, voice, mouth, eyes and/or head, and so on.

3.4 Lexicon and Prosody with regard to Emotions

3.4.1 Prosodic Levels for Emotions

Much research has been conducted on prosodic characteristics in utterances according to each emotion category. For example:

[2] The Table 2 comes from the working sheet simplified.

- Yamazaki [34] reports French subjects associate 'positive' emotions with F0 raising contour and the 'negative' emotions with F0 falling contour from her experiments of perceptual aspects of 'positive' or 'negative' emotions for synthetic stimuli.
- Halliday [35] points out a correlation between sentence types and prosodic manner, and claims a wh-question with a rising tone is 'tentative', while a yes/no question with a falling tone is 'peremptory'.
- Wichmann [36] mentions *High and Low* contour of please-request sentences can express a request of greater urgency.
- e.g. (High * Low) Please open the (Low*High)door.
- She concludes that *affective meanings are conveyed not only by continuously variable phonetic parameters, nor only through iconic associations with relative pitch height, but also by the conjunction between categorical choices of contour and utterance type.*
- Cælen-Haumont [37] defined 'melism' as a notion characterizing acoustic modifications of F0 (a great amplitude of F0), related to the expression of a linguistic meaning in affective conditions of speaking, and developed two models which she integrated in the Praat speech analysis software to measure the 'melism' [38].
- She confirmed by testing the models that the application of 'melism' to semantic or pragmatic analysis of utterances is possible.
- Aubergé shows that *prosody is one of the medium to express emotions in speech, through an in voluntary control* [39] by measuring acoustic parameters for strategic smile and spontaneous smile.
- Amir and al. extract an acoustic feature set of 12 elements from their corpus evaluation, which enables to classify emotional contents in speech: pitch and intensity statistics [44].

Thus we can confirm that information of prosody in utterances is indispensable for automatic recognition of emotions as well as lexemes and utterance types.

3.4.2 Manual Annotation for Emotions

The subjects of telephone conversations recorded at a French university are room reservation, schedule arrangement, taking a message, order of office supplies, etc., and some chats also are contained.

In the telephone messages at a public hospital, callers complain about problems with their computers or the software they use, and ask for technical help from an engineer, or ask for a rapid validation of an electronic access card for newcomers. In this con- text, a typical lexicon, or set of phrases expressing irritation, uncertainty or hesitation appear in the messages: *pénible, très pénible, drôlement embêté, une catastrophe, désespéré, relativement énervant, Ça me dérange beaucoup, ceci est assez désespérant, c'est embêtant,* etc. There are also "*C'est vraiment très urgent, Pourriez-vous venir voir?, Si vous pouviez passer rapidement"* etc. as more context-dependent examples.

In Table 1, we illustrate expressions of different emotions used in the telephone messages and the conversations. The first column shows emotion types, and the second and third column contain cited examples for the indicated emotion type.

Table 1. Lexemes having emotional content.

emotions	lexicon in the telephone messages	lexemes in the telephone conv.
happiness	merci beaucoup bonne journée,	Ah chouette!, Ouais!, impeccable!, Y a pas de souci!, C'est gentil. Merci!
disgust	nous avons un ordinateur qui est foutu où on travaille beaucoup,	
fear	J'ai peur que, je crains que,	J'ai peur que, je crains que
irritation	C'est embêtant, on est drôlement embêtés, ça me dérange, ça pose un réel problème, c'est relativement énervant, C'est très pénible, ma carte professionnelle de santé ne fonctionne plus, enfin, ça me marque 'illisible' ah !,	Ah non!
hesitation uncertainty	Je ne sais que faire, comment faire, nous aimerions savoir, je ne sais plus quoi faire. Euh, mon outlook ne s'ouvre plus, euh, et auparavant, il y a un message qui dit qu'il y a un problème avec le lecteur, on m'a dit qu'il y avait un problème sur... c'est que c'était plein, mon dossier était plein ?	Voyons voir, attends… je regarde, ben ben…, attends voir; heuh heuh, heueueueuh, Bof bof bof, hum hum, je ne sais pas, je vois pas bien, ça va faire un peu juste; on sait pas
sadness/ disappointment	Ici infirmière en état désespéré, j'ai mon poste qui est bloqué, on arrive pas non plus à arrêter, et ça c'est depuis hier, quoi,	Oh la pauvre! ah mince, ah zut, c'est dommage
surprise	Nous avons un écran noir ! nous pouvons ni voir les résultats, ni faire les mouvements à ce jour !	Oh, ça alors, ah bon?, tu crois?
anger	Les portables ne fonctionnent plus, nous sommes bloqués !, C'est urgent pour nous, alors je pense que ce soir, c'est plus envisageable, mais demain il faut impérativement, demain matin que ce problème doit être réglé.	Y en a marre!, C'est pas vrai!,
anxiety	C'est une catastrophe,	ça m'ennuie un peu, nous sommes ennuyés, c'est ennuyeux, y a une boulette, y a un souci, y a un truc qui me chiffonne,

Emotion eliciting words or phatics for each emotion class are surely found, but the prosodic features for each emotion class are divergent, as shown in Table 2., whereas there are clear prosodic signs which are confined to only one word which is semantically less significant. For example, in the utterance "*Il faudrait impérative-*

ment résoudre ce problème ce matin.", only '*matin*' is heavily accented, all of other words are uttered in a neutral tone, and we can interpret this accentuation as an implicit insistence on an urgent intervention.

We also have verified prosodic characteristics for some lexemes and set of phrases in the messages on Praat [14], but further examinations should be made to study the variety of prosodic features for each emotion class.

For instance, "*c'est relativement énervant*" or "*c'est très pénible.*" is uttered either with a neutral tone or at a raising tone.

Table 2. Lexical units and their prosody in Corpus Hotline CHRU

items	lexical units	emotional state	examples
same lexical unit with different prosodies	urgent	emphasized	on a un petit problème urgent
	extrême urgence	neutral	il nous le faut d'extrême urgence
	ce matin	emphasized	Il faudrait impérativement résoudre ce problème ce matin
		neutral	parce que ma collègue déjà appelé ce matin
lexemes having emotional contents	au secours, au secours	neutral	au secours, au secours, il faut absolument que je travaille sur cet ordi…
	c'est infernal	neutral	on passe des heures à connecter déconnecter l'ordinateur […] pour travailler, c'est infernal.
	énervant	neutral	c'est relativement énervant
	désagréable	neutral	ceci est assez désagréable, toutes les semaines
	urgent devoir vouloir	emphasized neutral neutral	c'est urgent ! vous deviez passer, mais nous avons pas de nouvelles de votre part. Veuillez me rappeler le plus tôt possible
utterance without any lexeme having emotional content	Il y a 9 étiquettes sur une et 16 étiquettes sur l'autre	irritated	L'imprimante nous imprime les étiquettes sur deux feuilles, qui sont toutes les deux incomplètes. Il y a 9 étiquettes sur une et 16 étiquettes sur l'autre
lexemes not having emotional contents	connecter and lundi matin	emphasized	Je n'arrive absolument plus à me connecter […] Est-ce que vous pourriez intervenir lundi matin?
	toutes les semaines	insistent	ceci est assez désagréable, toutes les semaines
	écran noir	emphasized	Nous avons un écran noir ! nous pouvons ni voir les résultats, ni faire les mouvements à ce jour !
anglicism	out	emphasized	Notre ordinateur est "out", et deviez passer, mais nous avons pas de nouvelles de votre part.

These phenomena are parameterized as emotion eliciting factors in the structures of attributes and its values in the emotion representation.

Choice of lexemes and prosody are a bench mark for detecting and identifying emotions as mentioned above. So, it's useful to mark lexemes with some labels in a dictionary used just like restricted UWs in UNL.

We propose, due to this fact, a set of emotion labels composed of 9 classes excluding "neutral" in our emotion classes and annotate lexemes in a UNL manner.

e.g.
désagréable (icl>sentiment>disgust)
catastrophe(icl>sentiment>anxiety)

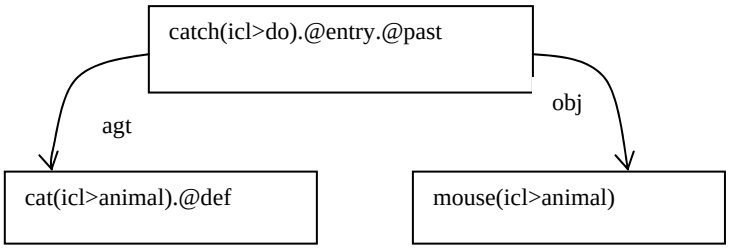
4 Semantic Representation of Emotions

4.1 UNL

A UNL graph consists of “UWs”, “Relations”, “Attributes”. It can be represented using tags. The “UWs” form the vocabulary of the UNL language, and denote "interlingual acceptions" (word senses). “Relations” and “Attributes” mainly make up the syntax, and the "knowledge base" (KB) covers the semantics of UNL [4]: it is a network containing all UWs, with all possible binary "semantic" and "thesaurus" relations between them. The 41 semantic relations contain "volitive agent, "coagent", "deep object", "instrument", "place", "place to", "time", "reason", "scene", "apposition", etc. Thesaurus relations contain "part of", "synonym", "is a", "field", and "antonym".

Here is an example of UNL graph, with one of its linear writings.

e.g. The cat caught a mouse.
[S] agt(catch(icl>do).@entry.@past, cat(icl>animal).@def)
obj(catch(icl>do).@entry.@past, mouse(icl>animal)) [/S]



An UW is made up of a character string followed by a list of constraints. UWs include basic UWs (bare English words), restricted UWs (English words with a constraint list), and extra UWs, which are a special type of restricted UWs [4].

One of the main advantages of UNL is the Universal Word (UW) dictionary, which enables us to specify word meanings at a deep level and to perform lexical disambiguation in a semantic oriented formalism.

In the example, “icl” in the constraint list enables us to define a subconcept of a basic UW. We will also apply later this constraint way to lexemes having emotional content for the purposes of representing emotions.

“agt” and “obj” are Relation tags, which indicate dependency relations between a head word in linguistic categories and other words, based on a case grammar type specification. “.@entry”, “.@past” and “.@def” are called Attribute tags, which indicate the grammatical conditions of a given utterance. The graph in the example does not contain any tags apart from the UNL tags, so they will be merged with other tags added as well as embedded in another format for the purposes of SSMT.

4.2 Semantic Representation of Emotions within UNL

The UNL semantic representation for written texts is actually designed by a set of 111 tags, which are divided into the 41 Relation tags and the 70 Attribute tags [4]. As for SSMT, some tags covering spoken language properties are merged with the UNL tag set: the added tags are 9 emotion tags we proposed, 8 prosody tags coming from the W3C recommendation [15], 12 behaviour tags from MPEG-4 [16] and, in particular, 28 speech act tags from a speech act research team [17] [18], and 5 interaction manner tags from GDA [19].

The UNL representation is a graph and consequently is not easy to encode in a linear data stream[3]. However, it is feasible to project it onto a description format such as XML, which authorizes the definition of elements and attributes. The representation obtained offers the same expressive power as graphs, but in the form of tags, and is easy to transmit. It is therefore easily interpreted by a DTD (Document Type Definition) conforming to the XML norm [26]. Thus, we attempted to transform UNL graphs into XML format as it facilitates speech synthesis information after the generation of the target language.

The representation schema of emotions proposed is made by adding tags expressing emotions according to the UNL. There are three ways to add such tags, that’s adding tags: “outside” of UNL makers as <VoiceUNL>, “inside” UNL text or a combination of both [21]. When emotions are formalized “inside” of the UNL makers, all tags representing prosody, behaviour and the speech act one are put in an UW. Therefore, in UNL graphs the arc concept representing a semantic relationship between two UWs might turn out to be unclear.

On the other hand, when emotions are formalized “outside” the UNL marker, in order to synchronize character’s strings and speech and visual items occurring simultaneously in an utterance, the same character’s string appears several times in a semantic representation. In such a dilemma, we create an additional UW type, which enables us to link speech, gesture, emotion and prosody tags: SP01, SP02, SP03...., and we use them in an “outside” and “combined” manner.

The following is a representation of an exchange in the “combined” manner:

[3] On Ariane-G5, which is an environment of MT into French language, UNL graphs are converted into tree structures [42].

```
<?xml version="1.0" encoding="iso-8859-1 ?>
<D dn=" TV " on="mt" dt="2003">
<Paragraph number="1">
<Sentence  snumber="2">
<org lang="el"> No! you may not drink, Victor.[4]</org>
<unlsem>
agt:SP01(drink(icl>do).@entry.@present.@obligation-not,
you.@emphasis)
mod:SP02(drink(icl>do).@entry.@present.@obligation-
not,no(icl>sentiment>surprise).@emphasis)
mod:SP03(no(icl>sentiment>surprise).@emphasis,
!(icl>symbol>surprise).@surprised)
mod:SP04(you,Victor(icl>name).@vocative)
</unlsem>
<VoiceUNL><speech-act>type="inform"    mod:SP01,    type="No"
agt:SP02 </speech-act>
<interaction> ref="drink" agt:SP01 </interaction>.
<emotion> class="surprise" mod:SP02 </emotion>
<gesture>eyebrows="left-and-right-raised" mod:SP02 </gesture>
</VoiceUNL>
</Sentence >
</Paragraph>
</D>
```

Note that "no" is annotated as "no(icl>sentiment>surprise)" by one of emotion class tags. It means that this "no" refers to a surprise as well as to a negation[5].

On the other hand, prosody tags (.@emphasis) are attached on UWs between <unlsem> and </unlsem>, and the gesture, emotion and discourse tags are external to <unlsem>, because only prosody is identified at the level of UWs, and the rest is often associated with utterance fragments or an entire utterance.

"drink", "you", "no", etc. are pivot languages called UW, and are converted into "boire", "tu", "non" respectively in the French generation module [2] [20][23]. Therefore, the transcription of this utterance is: “Non! tu ne peux pas boire, Victor”.

5 Conclusion

Our emotion studies have shown emotion analysis and generation were important issue in SSMT and our dialogue corpus analysis has suggested that lexemes are the most eleciting elements of emotions and that there is a delicate relationship between the lexeme uttered and its prosody. Thus we have proposed a semantic representation of emotions where all emotional expressions such as lexemes, prosody, gestures, etc. are described at the same time, by annotating lexemes with a set of labels, and adding speech property tags, speech act tags, interaction manner tags and behaviour tags to UNL in order to suit it to SSMT.

---

[4] The example cited here is: “May I drink?” “No!, you may not, Victor”. We have minimally changed it for convenience’s sake.

[5] Many previous studies have indicated that F0 raising contour is evoked by the happiness, surprise and anger in contrast to F0 falling contour which is evoked by the sadness or the uncertainty [3, 22]. This "no" is uttered in strong raising F0 contour on Praat.

We have found overlapping of utterances, irregular turn taking, category omission, deictic expressions, discourse ellipsis, etc. in our corpus [17]. Such interaction manners also are concerned with emotions of the speaker.

We actually use 5 tags from GDA[6] tag set [19] in the same way as paralinguistic tags to represent them as specificity of oral interaction manners. However further reflection is needed for discourse processing. For example, "VoiceXML" recommended by W3C [43] is designed for generating audio dialogues in monolingualism mainly on man-machine interactive system. A voice XML document is composed of top-level elements called "dialogs", and there are two types of "dialogs" : "forms" and "menus". "Forms" present information and gather input according to "Form Interpretation Algorithm", and "menus" offer choices of what to do next by referring one or more grammars associated with "dialogs". We also might need a mechanism which enables constantly to watch a flow of dialogues for coping with discourse ellipsis, anapholic expressions, etc. in oral dialogues.

The next step will be to develop a prototype with a speech and image interface as well as to enrich our corpus with speech and sound.

Acknowledgements The authors thank Professor Ch.Boitet, head of GETA in the CLIPS lab, for many inspiring discussions and the opportunity to conduct this research in the framework of the "UNL" project.

References

1 Chollet G., *Lingtour, Project Brochure*, GET, Paris, 2003.
2 Blanc E., GETA, CLIPS, IMAG, French deconverter, 2000.
3 Scherer, K. R., *Psychological models of emotion*, in J. Borod, *The neuropsychology of emotion*, Oxford/NewYork University Press, 2000.
4 UNL Center and UNL Foundation, *The Universal Networking Language (UNL) Specification*, Version 3, edition 1, UNU, IAS, Tokyo, Japan, 2002.
5 TV programmes on the 5th of October, 'Victor', Arte, 2003.
6 Holzman Lars E. and Pottenger William M., *Classification of Emotions in Internet Chat: An Application of Machine Learning Using Speech Phonemes*, 2003.
7 Fitrianie S., Wiggers P., and Leon J.M.,Rothkrantz L., *A Multi-Modal Using Natural Language Processing and Emotion Recognition*, Text, Speech and Dialogues, 6th International conference, TDS 2003, Czech Republic, Proceedings, LNAI 2807, V. Matousek and P.Mautner (Eds), Springer, 2003.
8 Randall, R.D., *The nature and Structure of Emotions*, U.S. Military Academy, USA, 1998, http://neologic.net/rd/Papers/EM-DEF19.html
9 Ortony A., Clore G. L, Collins A, *The Cognitive Structure of Emotions*, Cambridge, 1990.

[6] The Global Document Annotation (GDA) Initiative research team has proposed (2001)a XML-based tag set to help computing machines automatically infer the underlying semantic/pragmatic structure of documents. The GDA tag set is designed so that the GDA-annotation reduces the ambiguity in mapping a document to a sort of entity-relation graph (or semantic network) representing the underlying semantic structure [19]. There is a mapping schema between UNL specification tags and GDA one.

10 Plutchik R., The nature of emotions, American Scientist 89 344, USA, 2001.
11 Ekman P., *Emotions Revealed: Recognizing Faces and Feelings to Improve Communication and Emotional Life*, Times Book, USA, 2003.
12 Corpus DialAdmin, document CLIPS, Grenoble, 2003.
13 Corpus Hotline CHRU, document CLIPS, Grenoble, 2003.
14 Boersma P. and Weenink D., Praat, *Doing Phonetics by Computer, (version 4.1)*, Institute of Phonetic Sciences, University of Amsterdam, http://www.fon.hum.uva.nl/praat/, 2003.
15 W3C, *Speech Synthesis Markup Language Version 1.0*, W3C Working Draft 02, 2002, http://w3.org/TR/2002/WD-speech-synthesis-20021202, W3C, *Semantic Interpretation for Speech Recognition*, W3C Working Draft 02, 2003, http://w3.org/TR/2003/WD-semantic-interpretation-20030401
16 Koenen P. R. (ed), MPEG-4 Overview, http://mpeg.telecomitalialab.-com/standards/mpeg-4.html, 1999.
17 Tomokiyo M., Analyse discursive de dialogues oraux en français, japonais et anglais, Septentrion, Lille, 2001
18 Seligman M., Tomokiyo M. and Fais L., *A Bilingual Set of Communicative Acts Labels for Spontaneous Dialogues*, Rap. ATR, TR-IT-161, Japan, 1996.
19 Hasida K., *The GDA Tag Set (version 0.68)*, http://i-content.org/gda/tagman.html (version 0,71), 2003.
20 Southern Kings Consolidated School, *Types of emotions*, 2003 http://www.edu.pe.ca/southernkings/emotionstudi.html
21 Tomokiyo M. and Chollet G., *VoiceUNL: a proposal to represent speech control mechanisms within the Universal Networking Digital Language*, International Conference on the Convergence of Knowledge, Culture, Language and Information Technologies, Egypt, 2003.
22 Caelen-Haumont, G., BEL, B. *Subjectivité et émotion dans la prosodie de parole et du chat : espace, coordonnées et paramètres*. Colloque international « Emotions, Interactions & Développement », Grenoble, 2001. In Coletta, Jean-Marc; Tcherkassof, Anna (eds.) Perspectives actuelles sur les émotions. Cognition, langage et développement. Mardaga: Hayen. 2002.
23 Sérasset G., Boitet, Ch., *UNL-French deconversion as transfer & generation from an interlingua with possible quality enhancement through offline human interaction*, MT-SUMMIT VII, Singapore, 1999
24 HUMAINE, *HUMAINE Technical Annex*, 2004, http://emotion-research.net/
25 Corpus TRANSPORT: *Basic conversations on transport in English, French, Japanese and Chinese*, CLIPS, 2004.
26 Tsai WJ., *UNL news*, http://www-clips.imag.fr/geta/User/wang-ju.tsai/showunl.html, 2003.
27 Cowie R., *Describing the emotional states expressed in speech*, Proceedings of the ISCA Workshop on Speech and Emotion, Belfast, Proceedings on line, http://www.qbc.ac.uk /en/isca/proceedings, 2000
28 Cornelius R.R, *Theoretical approach to emotion*, Proceedings of the ISCA Workshop on Speech and Emotion, Belfast, Proceedings on line, http://www.qbc.ac.uk /en/isca/proceedings, 2000
29 Levenson R.W., Ekman, P., Friesen, W.V., *Voluntary facial action generates emotion-specific automatic nervous system activity*, Psychophysiology, 27, (4) 363-384, 1990, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi-?cmd=Retrieve&db=PubMed&list_uids=2236440&dopt=Abstract
30 James, W., What is an emotion?, Classics in the History of Psychology, an Internet resource developed by Green, Ch. D., 1884, http://psychclassics.-yorku.ca/James/emotion.htm

31 Arnold, M. B., *Emotion and personality*, vol1, Psychological aspects, New York, ed Columbia University Press, 1960.
32 Handbook of Emotions, Edited by Lewis, M., Haviland-Jones, J. M., Guiford, USA, 2004,
33 Averill, J.R., *A constructivist view of emotion*, Emotion : Theory, research and experience, vol.1, New York Academic Press, edited by Plutchik, R. and Kellerman, H., 1980.
34 Yamazaki H., *Perception des émotions "positive'' et ''négative'' chez les auditeurs français et japonais à travers le contour de F0*, Proceedings on Webs of JEP2004, 2004 http://aune.lpl.univ-aix.fr:16080/jep-taln04/proceed/actes/jep2004/Yamasaki.pdf
35 Haliday M.A.K., *An Introduction to Functional Grammar*, London, Edward Arnold, 1994.
36 Wichmann A., *Attitudinal Intonation and the Inferential Process*, International Conference Speech Prosody 2002, 2002 http://www.lpl.univ-aix.fr/projects/aix02/sp2002/pdf/wichmann.pdf
37 Cælen-Haumont G., Auran C., *The phonology of Melodic Prominence: the Structure of melisms*, In the "2nd International Conference on Speech Prosody Proceedings, SP 2004, http://www.isca-speech.org/archive/sp2004/sp04_143.pdf
38 Cælen-Haumont G., *Valeurs pragmatiques de la proéminence prosodique lexicale : de l'outil vers l'analyse*, Proceedings on Webs of JEP2004, 2004, http://aune.lpl.univ-aix.fr:16080/jep-taln04/proceed/actes/jep2004/Cælen-Haumont.pdf
39 Aubergé V., *The Prosody of Smile*, ISCA Workshop on Speech and Emotion ; A conceptual framewor for research, Proceedings on line, 2000 http://www.qub.ac.uk/en/isca/proceedings
40 Scherer, K. R., *Toward a concept of "modal emotions"*. In P. Ekman & R. J. Davidson (Eds.), The nature of emotion: Fundamental questions (pp. 25-31). New York/Oxford: Oxford University Press, 1994.
41 Morita Y., *Japanese dictionary – adjectives and adverbs*, Koudansya, Japan, 1989.
42 Blanc E., *From the UNL hypergraph to GETA's multilevel tree.* Proceeding of Machine Translation, University of Exeter, 18-21 oct., pp9.1—9.9. Ed. British Comp. Society, 2000.
43 *Voice Extensible Markup Language (VoiceXML) Version 2.0*, W3C Recommendation, 2004, http://www.w3.org/TR/voicexml2
44 Amir N., Ron S., Laor N., *Analysis of an emotional speech corpus in Hebrew based on objective criteria*, ISCA Workshop on Speech and Emotion ; A conceptual framewor for research, Proceedings on line, 2000. http://www.qub.ac.uk/en/isca/proceedings