Greek Verbs and User Friendliness in the Speech Recognition and the Speech Production Module of Dialog Systems for the Broad Public

Christina Alexandris[1] and Ioanna Malagardi[2]

[1] National and Kapodistrian University of Athens, Greece
[2] Department of Informatics & Telecommunications
Educational & Language Technology Laboratory
National and Kapodistrian University of Athens
calexandris@gs.uoa.gr, imalagar@di.uoa.gr

Abstract. It has been observed that for the Greek language, in Service- Oriented Dialog Systems targeted towards the broad public, verbs display particular features to be considered both in the System's Speech Recognition (ASR) and Speech Synthesis (or Speech Production) Module. Additionally, the function of verbs, both in respect to their role in the semantic content of the spoken utterance and in respect to their prosodic features in the spoken utterance, is observed to be directly linked to the step and the related Speech Act in the dialog structure. From a prosodic aspect, it is observed that, in spoken input, "Multitasking" verbs do not receive prosodic emphasis, whereas in spoken output, prosodic emphasis is given on words signalizing the User-System Relationship.

Keywords: prosodic emphasis, verb ontology, User-friendliness, "Multitasking" verbs, User-System Relationship expressions.

1 Introduction

Service- Oriented Dialog Systems targeted towards the broad public involve a higher percentage of non-sublanguage specific vocabulary and a lower percentage of terminology and professional jargon. Unlike highly specialized Task-related Dialog Systems, in Service- Oriented Dialog Systems, the Human-Computer interaction taking place is directed towards two equally significant goals, namely the successful performance of the activated or requested task and User satisfaction and User-friendliness. These goals are related to requirements on the Satisfaction Level in respect to a System's evaluation criteria, namely perceived task success, comparability of human partner and trustworthiness [10]. In some cases of successful design and implementation, even a sense of attachment of the User in respect to the System can be possible [9]. It should be noted that the more goals to be achieved, the more parameters in the System Design and System Requirements, and subsequently Dialog Design are to be considered [14]. Prosodic modelling for the Speech Acts in Service-Oriented Dialog Systems may, therefore, be characterized as a complex task.

C. Stephanidis (Ed.): Universal Access in HCI, Part IV, HCII 2011, LNCS 6768, pp. 3–11, 2011.

In User input recognized by Service- Oriented Dialog Systems, verbs play an essential role in the vocabulary of User input. In Service- Oriented Dialog Systems, verbs play an equally essential role when they are produced as spoken output by the System's Speech Processing Module. It has been observed that for the Greek language, in Service- Oriented Dialog Systems targeted towards the broad public, verbs display particular features to be considered both in the System's Speech Recognition (ASR) and Speech Synthesis (or Speech Production) Module.

The function of verbs, both in respect to their role in the semantic content of the spoken utterance and in respect to their prosodic features in the spoken utterance, is observed to be directly linked to the step and the related Speech Act in the dialog structure. In spoken Greek, non-sublanguage specific verbs in user input may be characterized by polysemy and absence of prosodic emphasis (I). In contrary, non-sublanguage specific verbs in the System's spoken output contain a semantic feature expressing the System-User relationship and receive prosodic emphasis for the achievement of friendliness towards the User (II). Furthermore, in User input, the correct interpretation of non-sublanguage specific verbs in user input requires the identification of the correct Speech Act (A). Similarly, in the System's spoken output, the context of a so-called "Non-Task-related" Speech Act in dialog structure [1] is the condition for non-sublanguage specific verbs to express the System-User relationship (B).

1.1 Verb Classes and Ontologies: Automatic Grouping Practices

Here, we will address two types of verb classes: (1) "Multitasking" verbs, usually occurring high in the verb ontology, some of them constituting "primitive" verbs and (2) verbs constituting "User-System-Relationship" expressions, usually occurring lower in the verb ontology, most of them constituting verbs with one or more objects.

The "Multitasking" verbs used in input phrases must be interpreted in their basic "primitive" interpretation and the verbs constituting "User-System-Relationship" expressions, usually occurring lower in the verb ontology, are interpreted in the form they occur. The analysis of the definitions of the above-mentioned classes of verbs is used for the application of the understanding and description of actions.

Verb classes can be identified and grouped automatically. Previous studies [5], [8] have demonstrated that computational processing of verbs involves a process of automatic grouping of verbs, namely that the definitions of the verb's features are given as input to a System producing an output that renders a grouping of the verbs and synthesized definitions of these verbs using "primitives". The set of verb entries related to the above categories used as input to a System can be identified on the basis of chains containing their definitions. The verb at the end of a chain was used as the criterion of verb identification and verb grouping [5], [8].

Here, we note that prior to using these chains, it was necessary to eliminate the cyclic parts of the definition chains which were also automatically detected by the System. The definitions of the verbs in each definition are, in turn, retrieved from the lexicon and, in this way, chains of definitions are formed. These chains end up in circularities that correspond to reaching basic verbs, "basic", representing primitive actions [5]. The elimination of the circularity that occurs in certain chains requires the choice of the suitable verb as a terminal of the chain. The choice for each case of elimination of circularity requires the adoption of some ontology [5], [8].

1.2 Verb Classes and Ontologies – The Prosody and Semantics Relation

In respect to Speech Processing, "Multitasking" verbs are observed to be related to the special prosodic feature of the absence of prosodic emphasis in regard to the recognition of the Speaker's or User's spoken input. In contrary, verbs constituting "User-System-Relationship" expressions are observed to be related to the prosodic feature of prosodic emphasis in respect to the production of the Systems's spoken output. In respect to User-friendliness, within the Speech-Act framework of Service-Oriented Dialog Systems, the absence of prosodic emphasis on "Multitasking" verbs contributes to the achievement of clarity and processability of User-input, while the possibility of using these verbs in User input allows the User to speak in a more natural way to the System, without too many restrictions in utterance content. On the other hand, prosodic emphasis on "User-System-Relationship" verbs contributes to the achievement of naturally sounding, comprehensible and friendly System-output.

The relationship of "Multitasking" and "User-System-Relationship" verbs in respect to prosodic emphasis may be depicted in the following table (Table 1).

Table 1. Relationship of "Multitasking" and "User-System-Relationship" verbs in respect to prosodic emphasis

[-prosodic emphasis]:	[+prosodic emphasis]:
System Input	System Output
Verb type: "Multitasking" verbs	Verb type: "User-System-Relationship" expressions
Speech Act: Task-Related	Speech Act: Non Task-Related
User-friendliness Goal: clarity and processability for System with few restrictions in User-Input	User-friendliness Goal: naturally sounding, comprehensible and friendly System-Output

Furthermore, it may noted that, at least within the Speech-Act framework of Service-Oriented Dialog Systems, verbs with a relatively high content of specifying features in their semantics, such as verbs signalizing the User-System-Relationship, have a higher likelihood to receive prosodic emphasis, whereas verbs with a more vague or general semantic content and often occurring higher in semantic ontologies ("primitives") have a higher likelihood not to receive prosodic emphasis. The question of whether similar observations can be made in respect to other verb classes and/or contexts other than the Speech-Act framework of Service-Oriented Dialog Systems, remains an issue under investigation.

2 Prosody and "Multitasking Verbs" in Spoken Input

From the aspect of the System's Speech Recognition (ASR) Module, one basic problem that requires to be addressed is the existence of "Multitasking" verbs which may be described as verbs related to multiple semantic meanings and used in a variety of expressions, existing in Modern Greek, and possibly in other languages as well. For

example, in the sublanguage related to the communication context of the Citizen-Shield system for the Service sector (consumer complaints) [11], the semantically related verbs "buy", "get" and "purchase" may be used in similar expressions as the (primitive) verbs "is" and "have", as well as the verbs "give" and "receive", to convey the same semantic meaning from the speaker. This possibility is illustrated in the following examples (Table 2).

Table 2. "Multitasking" verbs in the communication context of the CitizenShield system for the Service sector (consumer complaints) (User-input)

[-prosodic emphasis]	Examples (User-input):
have, get	"I got this cereal" / "I have this cereal here" "They give misguiding information on the package"
be, is	"They are from a convenience store"
	"This is about a cereal"
see	"I saw some misguiding information on the package"

The rather rigid and controlled nature of the CitizenShield System's dialog structure allows the application of constraints in regard to the processing of the correct interpretation of "Multitasking" verbs. The CitizenShield System's dialog structure is based on Task-related Speech Acts involving Directed Dialogs [15], [16], most of which involve Yes-No Questions or questions directed towards Keyword answers. From a prosodic aspect, it is observed that, unlike sublanguage-specific verbs with more restricted semantics, "Multitasking" verbs do not receive prosodic emphasis when uttered by Greek native speakers as spoken input to be processed by the System's Speech Recognition (ASR) Module and are, therefore, of less significance in the semantics of the spoken utterance. This feature enables their efficient processing, since they are recognizable by the absence of prosodic emphasis and, consequently, can be identifiable by an attached marker signifying absence of prosodic emphasis. For the achievement of clarity and processability as spoken input to be processed by the System's Speech Recognition (ASR) Module, "Multitasking" verbs do not receive prosodic emphasis and, therefore, an additional prosodic marker signifying absence of prosodic emphasis is proposed.

Both in monolingual and in multilingual applications, the traditional approach constructed around a verb-predicate signalizing the basic semantic content of the utterance, the so-called "frame", also typically used in traditional Interlinguas [6], may not always be adequate for the polysemy of "Multitasking" verbs in User input. Thus lexical-based alternatives linked to Speech Act types in respect to the steps dialog context are proposed. Specifically, the correct interpretation of the semantics of "Multitasking" verbs is directly related to the type of Speech Act expressed in the spoken utterance.

In an attempt to meet the needs of the diverse community of foreign residents in Greece, Interlinguas (ILTS) are used as semantic templates for a possible multilingual extension of the CitizenShield dialog system for consumer complaints. The proposed Interlinguas are designed to function within a very restricted sublanguage related to a specific task, but at the same time, for a very diverse range of languages and language families, allow minimum interference of language-specific factors.

Table 3. Examples of Simple Interlinguas with "Multitasking" verbs

S-FRAME	Keyword and Lexical Content of "S-ILT"
[S-FRAME = GOT [-prosodic emphasis]	WHO (PERSON) WHAT (ITEM) QUANTITY (QUANT), PRICE (NUM-EURO) WHERE (PLACE) WHEN (DAY)]
[S-FRAME = HAVE [-prosodic emphasis]	WHO (PERSON) WHAT (ITEM) QUANTITY (QUANT), PRICE (NUM-EURO) WHERE (PLACE) WHEN (DAY)]
[S-FRAME = IS [-prosodic emphasis]	WHO (PERSON) WHAT (ITEM) QUANTITY (QUANT), PRICE (NUM-EURO) WHERE (PLACE) WHEN (DAY)]

The proposed S-Interlinguas (Table 3) may be characterized by a very simple structure and to be more of Interlinguas with an accepting or rejecting input function rather than the traditional Interlinguas with the function of summarizing the semantic content of a spoken utterance [6], [12]. Thus, in the present application, the role of the "frame" in the S-Interlingua structure is weakened and the core of the semantic content is shifted to the lower level of the lexical entries. The lexical level allows the possibility to add a large number of variants of terms in different languages.

The simple structure of the S-Interlingua (S-ILT) may be differentiated in three levels: (1) the Template Level, the Keyword-Content Level (2) and (3) the Lexical-Level. The Template Level constitutes the top level and is loosely associated with the verb of the speakers utterance, however, its function is more that of "wrapping up" the key-word contents of the speakers response than the original function of signalizing the sentence content. The "frame" level (S-FRAME) will not signalize the meaning of the sentence: this task will be performed by the lexical entries. However, in most cases, the "frame" level is retained for purely functional purposes, namely for connecting the lexical entries to each other and facilitating the rejection of possible input whose semantic content is irrelevant and not contained within the sublanguage of the application. This rejected input, however, may contain individual words initially recognized by the system at the Lexical Level

3 Prosody and "System-User Relationship" Verbs in Spoken Output

In respect to spoken output by the System's Speech Processing Module, it has been observed that for languages like Greek, where friendliness is related to directness and spontaneity, constituting features of Positive Politeness [13], from a prosodic aspect, User-friendliness can be achieved with prosodic emphasis on verbs functioning as elements expressing the User-System Relationship.

Words signalizing the User-System Relationship can be subsumed under the general category of expressions involving the System's or User's positive intention or cooperation and may be related to respective Non-Task-related Speech Acts. These word categories may be described as expressions related to the System's positive attitude toward the User and be categorized as (1) system-service verbs, (2) system-intention verbs, (3) system-service nouns, (4) system-intention nouns, (5) user-intention verbs and (5) user-action verbs.Typical examples of system-service verbs are the verbs "give" and "show" and nominalized user-intention verbs such as "cooperation". Examples of system-intention verbs (including nominalized system-intention verbs) are the verbs "help" and "assist". The verbs "wish"/"want" and "finished" are typical examples of user-intention (or user intended action) and user action verbs respectively.

Specifically, User-friendliness can be achieved with prosodic emphasis on verbs (or nominalized verbs) functioning as elements expressing the User-System Relationship in the context of Non-Task-related Speech Acts such as "Apologize" and "Delay" [1].

3.1 Defining Non-task Related Speech Acts in Dialog Structure

Data from European Union projects in Speech Technology for social services and Human-Computer Interaction in English, German and Greek [17],[18],[19],[20] allows the formulation of a general categorization scheme of Non-Task-related Speech Acts. Specifically, Non-Task-related Speech Acts can be divided into three main categories: Speech Acts constituting an independent step in dialog structure (Category 1, for example, "Close Dialog":"Thank you for using the Quick-Serve Interface"), Speech Acts attached to other Task-Related Speech Acts [4] constituting with them a singular step in dialog structure (Category 2, for example, "I am sorry" ("Apologize") following or preceeding the Task-Related Speech Act ("Inform"), "Your input cannot be processed" or "I cannot understand your request" ("Justify") following or preceeding the Task-Related Speech Act ("Request")) and Speech Acts constituting an optional step in dialog structure of Service-Oriented dialogs (Category 3), for example, "Reminder": "You still have two minutes to complete this transaction" [1].

3.2 Prosodic Modeling and the User-System Relationship in Greek

In the following examples (Table 4) from the CitizenShield dialog system [11], the above listed types of words signalizing the User-System Relationship receive prosodic emphasis ("Usr-Sys-Rel prosodic emphasis", indicated as [Usr-Sys-Rel prosod] following the emphasized word). These words are the expressions "sorry" (system-intention noun – in Greek), "hear", "request" (system-service verbs), "thank" (system-intention verb), "cooperation" (nominalised user-intention verb) and "completed" (user-action verb).

At this point, it is important to stress that not all Non-Task-related Speech Acts necessarily contain "Usr-Sys-Rel" expressions. It should, additionally, be noted that expressions signalizing negations, temporal and spatial information, quantity and quality, as well as sublanguage-specific task-related expressions receive prosodic emphasis by default ("default prosodic emphasis", indicated in italics) [2]. In Table 4,

these are the expressions “not”, “some”, “very much”, “more”, “additional” and “obviously” [3].

Additionally, it should be stressed that in Non-Task-related Speech Acts, “Usr-Sys-Rel prosodic emphasis” has a priority over “default prosodic emphasis” in respect to amplitude. Specifically, in Non-Task-related Speech Acts, the amplitude of the prosodic emphasis on Usr-Sys-Rel expressions is intended to be slightly higher than the amplitude of the prosodic emphasis on expressions receiving default prosodic emphasis.

Table 4. Examples with User-System-Relationship Expressions (CitizenShield System)

[+prosodic emphasis]: for “Usr-Sys-Rel” expressions and “default” prosodic emphasis

1. **Συγγνώμη**[Usr-Sys-Rel prosod] ***δεν*** σας **άκουσα**[Usr-Sys-Rel prosod]
   (Non-Task-related Speech Act: “Justify”)

2. Θα σας **ζητήσω**[Usr-Sys-Rel prosod] ***μερικές*** πληροφορίες ***ακόμα***
   (Non-Task-related Speech Act: “Introduce-new-task”)

3. Σας **ευχαριστούμε**[Usr-Sys-Rel prosod] ***πολύ*** για την **συνεργασία**
   [Usr-Sys-Rel prosod]σας
   (Non-Task-related Speech Act: “Thank”)

4. ***Προφανώς*** **ολοκληρώσατε**[Usr-Sys-Rel prosod] με τις ***επιπλέον*** πληροφορίες
   (Non-Task-related Speech Act: “Reminder”)

Translations close to the syntax of original spoken utterances

1. I am **sorry**[Usr-Sys-Rel prosod]**,** I could ***not*** **hear**[Usr-Sys-Rel prosod] you**.**
   (Non-Task-related Speech Act:“Justify”)

2. I will **request** [Usr-Sys-Rel prosod] from you ***some more*** information.
   (Non-Task-related Speech Act: “Introduce-new-task”)

3. We **thank**[Usr-Sys-Rel prosod] **you** ***very much*** for your **cooperation**
   [Usr-Sys-Rel prosod]
   (Non-Task-related Speech Act: “Thank”)

4. You have ***obviously*** **completed**[Usr-Sys-Rel prosod] providing the
   additional input
   (Non-Task-related Speech Act: “Reminder”)

We note that, in Greek, as a verb-framed and pro-drop language (like Spanish or Italian), the prosodic emphasis is directly matched to the finite verb, containing the features of the verb’s subject - in this case the System or the User. This difference in respect to languages such as English may also influence the process of identifying Usr-Sys-Rel expressions.

For example, the Greek verb “zi´tiso” (“request”) in the context of “make questions” contains the features of the verb’s subject. In the context of Service-oriented HCI applications, the Greek verb “zi´tiso” may be identified as an Usr-Sys-Rel expression. In another example, the Greek verb “olokli´rosate” (“finished-completed”)

is equivalent to the verb "finished" in English. We note that the semantics of the Greek verbs "zi΄tiso" and "olokli΄rosate" allows them to be classified as a Usr-Sys-Rel expression, whereas the respective verbs "request" and "finished" in English are classified as verbs signalizing an ACTION-TYPE [7], a task-related expression receiving default emphasis [1],[3],[7].

4 Conclusions and Further Research

In Service- Oriented Dialog Systems, the semantic content and the prosodic features of verbs are observed to be directly linked to the step and the related Speech Act in the dialog structure.

In spoken Greek Input, non-sublanguage specific verbs in user input may be characterized by polysemy and absence of prosodic emphasis (1). In contrary, non-sublanguage specific verbs in the System's spoken output contain a semantic feature expressing the System-User relationship and receive prosodic emphasis for the achievement of User-friendliness (2).

Specifically, it is observed that, unlike sublanguage-specific verbs with more restricted semantics, "Multitasking" verbs do not receive prosodic emphasis when uttered by Greek native speakers as spoken input to be processed by the System's Speech Recognition (ASR) Module and, therefore, play a less important role in the semantics of the spoken utterance. Thus, traditional practices based on a verb-predicate signalizing the basic semantic content of the utterance, the so-called "frame", also typically used in traditional Interlinguas [6], may not always be efficient for the processing of "Multitasking" verbs in User input. Therefore, lexical-based alternatives linked to Speech Act types in dialog context are proposed, especially in the case of multilingual applications in diverse languages and language families.

On the other hand, in spoken Greek output, User-friendliness can be achieved with prosodic emphasis on verbs functioning as elements expressing the User-System Relationship in the context of Non-Task-related Speech Acts, as defined, based on data within the framework of European Union projects. Prosodic emphasis can be directly matched to the finite verb, constituting a User-System Relationship expression and containing the features of the verb's subject - in this case the System or the User. This is possible in a verb-framed and pro-drop language such as Greek, however, may be problematic in non verb-framed and non pro-drop languages.

Further research is required to determine whether similar phenomena are observed in respect to verbs or to other word categories in Service- Oriented Dialog Systems of other languages, possibly also in Dialog Systems of different domains.

References

1. Alexandris, C.: Speech Acts and Prosodic Modeling in Service-Oriented Dialog Systems. In: Computer Science Research and Technology. Nova Science Publishers, Hauppauge (2010)
2. Alexandris, C.: Show and Tell: Using Semantically Processable Prosodic Markers for Spatial Expressions in an HCI System for Consumer Complaints. In: Jacko, J.A. (ed.) HCI 2007. LNCS, vol. 4552, pp. 13–22. Springer, Heidelberg (2007)

3. Alexandris, C.: Word Category and Prosodic Emphasis in Dialog Modules of Speech Technology Applications. In: Botinis, A. (ed.) Proceedings of the 2nd ISCA Workshop on Experimental Linguistics, ExLing 2008, Athens, Greece, pp. 5–8 (August 2008)
4. Heeman, R., Byron, D., Allen, J.F.: Identifying Discourse Markers in Spoken Dialog. In: Proceedings of the AAAI Spring Symposium on Applying Machine Learning to Discourse Processing, Stanford (March 1998)
5. Kontos, J., Malagardi, I., Pegou, M.: Processing of Verb Definitions from Dictionaries. In: Proceedings of the 3rd International Conference in Greek Linguistics, Athens, pp. 954–961 (1997) (in Greek)
6. Levin, L., Gates, D., Lavie, A., Pianesi, F., Wallace, D., Watanabe, T., Woszczyna, M.: Evaluation of a Practical Interlingua for Task-Oriented Dialog. In: Proceedings of ANLP/NAACL 2000 Workshop on Applied Interlinguas, Seattle, WA (April 2000)
7. Malagardi, I., Alexandris, C.: Verb Processing in Spoken Commands for Household Security and Appliances. In: Stephanidis, C. (ed.) UAHCI 2009, Part II. LNCS, vol. 5615, pp. 92–99. Springer, Heidelberg (2009)
8. Malagardi, I. , Kontos, J.: Motion Verbs and Vision. In: Proceedings of the 8th Hellenic European Research on Computer Mathematics & its Applications Conference (HERCMA 2007), Athens (2007), http://www.aueb.gr/pympe/hercma/proceedings2007
9. Matsumoto, N., Ueda, H., Yamazaki, T., Murai, H.: Life with a Robot Companion: Video Analysis of 16-Days of Interaction with a Home Robot in a "Ubiquitous Home" Environment. In: Jacko, J.A. (ed.) HCI International 2009. LNCS, vol. 5611, pp. 341–350. Springer, Heidelberg (2009)
10. Moeller, S.: Quality of Telephone-Based Spoken Dialog Systems. Springer, New York (2005)
11. Nottas, M., Alexandris, C., Tsopanoglou, A., Bakamidis, S.: A Hybrid Approach to Dialog Input in the CitzenShield Dialog System for Consumer Complaints. In: Proceedings of HCI 2007, Beijing, Peoples Republic of China (2007)
12. Schultz, T., Alexander, D., Black, A., Peterson, K., Suebvisai, S., Waibel, A.: A Thai speech translation system for medical dialogs. In: Proceedings of the conference on Human Language Technologies (HLT-NAACL), Boston, MA, USA (2004)
13. Sifianou, M.: Discourse Analysis. An Introduction. Leader Books, Athens (2001)
14. Wiegers, K.E.: Software Requirements. Microsoft Press, Redmond (2005)
15. Williams, J.D., Witt, S.M.: A Comparison of Dialog Strategies for Call Routing. International Journal of Speech Technology 7(1), 9–24 (2004)
16. Williams, J.D., Poupart, P., Young, S.: Partially Observable Markov Decision Processes with Continuous Observations for Dialog Management. In: Proceedings of the 6th SigDial Workshop on Discourse and Dialog, Lisbon (September 2005)
17. The Agent-DYSL Project, http://www.agent-dysl.eu/
18. The HEARCOM Project, http://hearcom.eu/main.html
19. The ERMIS Project, http://www.image.ntua.gr/ermis/
20. The SOPRANO Project, http://www.soprano-ip.org/