Word category and prosodic emphasis in dialog modules of speech technology applications

Christina Alexandris
Department German, University of Athens, Greece

https://doi.org/10.36505/ExLing-2008/02/0002/000061

Abstract

The prosodic behavior of word categories in Modern Greek is observed in relation to two keyword-based Speech Technology applications, namely for a Speech Recognition system in the healthcare domain for task-oriented dialogs intended for senior citizens and for the prosodic modelling of utterances produced by a Conversational Agent in a dialog system for consumer complaints. Word categories are evaluated and classified in the two applications, in respect to the types of words contained in each category, their relation to keywords and their effect on the user.

Key words: prosodic emphasis, semantic content, keywords, user-friendliness

System requirements and evaluation

The behavior of word categories in Modern Greek is observed during the prosodic modelling of utterances recorded in a studio for the construction of the speech output produced by a Conversational Agent in a dialog system for consumer complaints (CitizenShield System, www.polias.gr), labelled "CSh System", and for a dialog-module of a speech-recognition in a system for Senior Citizens which activates and monitors house appliances and the general room environment (SOPRANO Project, www.soprano-ip.org), labelled "SPR System". The acceptable prosody to the target user-groups, constituting the general public in the CSh-System and senior citizens in the SPR Project, is modelled with the use of emphasis as prosodic marker, placed on the appropriate elements of the utterance for achieving user-friendliness in man-machine communication in the sense of "accuracy" and "directness" (Hausser 2006) as well as (B) a sense of control to the user for the CSh System (Alexandris 2007, Nottas et al. 2007) and (C) familiarity for the SPR System as an additional feature, according to the SPR Project's User Requirements. The evaluation of the prosodic markers in the SPR Project was primarily based on Interaction Aspects used for testing the performance of the CSh's System, namely (1) the Utterance, (2) Functional and (3) Satisfaction Levels (Moeller 2005) and also due both to the complexity of the CSh's System (a) user-group, a remarkable subset of which constitutes senior citizens, according to the CSh's Project User Requirements (Work-Package 1) and the (b) relatively positive results of the evaluation of the systems prosodic modelling in the Evaluation Phase (Work-Package 7) (of a

50-Test Users Group, 60% Men and 40% Women, Average Age: 35 – 45, 80%, evaluated the Interface as “Very natural-sounding”, and 50% evaluated the System as “Seems to inspire very high credibility”). The prosodic modeling of the output in the SPR Project was subjected to an internal evaluation before the execution of user-tests to Senior Citizens.

In the present study we attempt to provide a differentiation between specific word categories in which prosodic emphasis does not determine their semantic content (I) and word categories whose semantic content may be determined by prosodic emphasis (II). In the first case, the semantic interpretation of the entire phrase or sentence may be determined by the type of element receiving prosodic emphasis, but the semantic content of the emphasized element itself is not effected. Based on the empirical data obtained both from the recorded corpora (ASR Component) and the evaluation of the spoken utterances produced by the Conversational Agent (Nottas et al. 2007) in the CSh and SPR Projects we attempt to go a step further and provide (1) an integration of the results of previous studies (Alexandris et al. 2005, Alexandris 2007) and to proceed to a (2) differentiation between specific word categories in relation to prosodic emphasis and their semantic content.

Prosody, word categories and semantic content

The group of word categories whose semantic content may be determined by prosodic emphasis namely (1) spatial and temporal expressions, (2), a subgroup of quantifiers and numericals and (3) a sub-group of discourse particles identified as “politeness markers” (Alexandris and Fotinea, 2004) is classified as Category A or “Prosodically Determined” words. For spatial and temporal expressions, and for the subgroup of quantifiers and numericals, the presence of prosodic emphasis signalizes an indexical interpretation (“exactly”) as opposed to a vague (Schilder and Habel 2001), interpretation or a fixed expression (Alexandris et al. 2005), where in the latter cases, there is an absence of prosodic emphasis. For example, with prosodic emphasis there is an indexical interpretation of the spatial expression “'dipla” as “along” in the sentence “the crack was exactly along (parallel) to the band in the packaging” as opposed to its vague interpretation as “next-to” in the same sentence. The same is observed for the temporal expression “'oso” with its indexical interpretation as “for as long as” in the sentence “the array is created for as long as the loop is running” as opposed to its vague interpretation as “while” in the same sentence. Similarly, the numerical or quantificational expression “two” (“d'yo”) is used in its indexical and literal meaning when it receives prosodic emphasis in the sentence “wait for two minutes”, while, in the same sentence without prosodic emphasis, it is perceived as a fixed expression (“wait a moment”).

For discourse particles identified as "politeness markers", the absence of prosodic emphasis signalizes them as politeness markers, while with the presence of prosodic emphasis they only have the property of discourse particles. Thus, absence of prosodic emphasis in the discourse particles "Tell me ('pite mou)" and "Mabey" ('mipos) signalizes positive politeness and friendliness towards the User in the following utterances produced by the Conversational Agents for the CSh and SPR Systems respectively: Tell me ('pite mou), what is the product (Preferred utterance by CSh Users), Mabey ('mipos) you want me to check the kitchen? (SPR)

The group of word categories where prosodic emphasis may emphasize or intensify, but may not determine the semantic content, is classified as Category B or "Prosodically Sensitive" words. This group involves (1) adjectives expressing quality and (2) adverbs expressing mode perceptible to the senses, used in a literal, non-metaphorical way. For example, prosodic emphasis on the adjective "round" ("strogi 'lo"), in the sentence "It was in a round box" (CSh Project) signalizes the meaning "truly/par excellence" round". Similarily, prosodic emphasis on the adverb "upside down" ("an 'apoda") for example, in the sentence "I turned it upside down" (CSh Project) signalizes the meaning "completely upside down".

Both Category A and Category B type prosodic emphasis may be used in both the CSh and the SPR Projects, for the (a) correct interpretation of Speaker Input in the respective Automatic Speech Recognition (ASR) Modules and (b) for achieving user-friendliness in man-machine communication in the sense of "accuracy" and "directness" (Hausser, 2006) towards the user in the Conversational Agent's spoken output.

The rest of the word categories that are not effected by prosodic emphasis in respect to their semantic content are classified as Category C or "Prosodically Independent" words. The presence or absence of prosodic emphasis on words of Category C only effects the semantic interpretation of the entire phrase or sentence in which they belong. A significant percentage of these words are nouns or verbs and they may constitute sublanguage-specific keywords. Prosodic emphasis on keywords focuses on the basic content of the utterance, for example, whether it is an action in question, in the case of a verb, or a specific object in question, in the case of a noun. Prosodic emphasis on the word elements of Category C, words is sentence-dependent and highly sublanguage- and application-specific. Prosodic emphasis on elements of Category C in the CSh Project is used both for (a) determining the basic content of the Speaker's input, (b) for directing the Speaker's input towards a keyword-specific answer, as well as (c) for achieving accuracy and directness in the Conversational Agent's output. For example, in the sentence "Please tell us any additional information you wish about the product or about your transaction" (CSh), the keywords "additional", "product" and "transaction" receive prosodic emphasis for clarity towards the Users and simultaneously direct towards obtaining a respective keyword-specific answer, in this case "product-type" and "transaction-type". Similarly, a "Yes/No" Answer is requested with the use of prosodic emphasis either on "check" or on "thermostat" in the question "Shall I check the thermostat?" in the SPR Project.

Systematic use of prosodic emphasis features

In contrast to both A and B word categories, or "Prosodically Determined" and "Prosodically Sensitive" words, whose plus or minus (±) prosodic emphasis features can be systematically used in various Speech Technology Applications, including Text-to-Speech (TTS) and ASR, the prosodic modelling of Category C or "Prosodically Independent" words is highly sublanguage-dependent and application-specific.

References

Alexandris, C. 2007. Show and Tell: Using Semantically Processable Prosodic Markers for Spatial Expressions in an HCI System for Consumer Complaints. Human-Computer Interaction. HCI Intelligent Multimodal Interaction Environments. 4552/2007, 13-22.

Alexandris, C. Fotinea, S-E and Efthimiou, E. 2005. Emphasis as an Extra-Linguistic Marker for Resolving Spatial and Temporal Ambiguities in Machine Translation for a Speech-to-Speech System involving Greek. In Proc. of the 3rd Intern. Conference on Universal Access in Human-Computer Interaction (UAHCI 2005). Las Vegas, Nevada, USA.

Alexandris, C., Fotinea, S-E. 2004. Discourse Particles: Indicators of Positive and Non-Positive Politeness in the Discourse Structure of Dialog Systems for Modern Greek, International Journal for Language Data ProcessingSprache und Datenverarbeitung, 1-2/2004, 19-29.

Hausser, R. 2006. A Computational Model of Natural Language Communication, Interpretation, Inference and Production in Database Semantics. Berlin, Springer.

Moeller, S. 2005. Quality of Telephone-Based Spoken Dialogue Systems. New York, Springer.

Nottas, M., Alexandris, C, Tsopanoglou, A. and Bakamidis, S. 2007. A Hybrid Approach to Dialog Input in the CitzenShield Dialog System for Consumer Complaints. In Proc. Of Human-Computer Interaction (HCI) 2007. Beijing, China.

Schilder, F. and Habel, C. 2001. From Temporal Expressions to Temporal Information. In Proc. of ACL-2001, Workshop on Temporal and Spatial Information Processing, 1309-1316, Pittsburgh, Pennsylvania, USA.