

"Show and Tell": Using Semantically Processable Prosodic Markers for Spatial Expressions in an HCI System for Consumer Complaints

Christina Alexandris

Institute for Language and Speech Processing (ILSP)
Artemidos 6 & Epidavrou, GR-15125 Athens, Greece
calex@ilsp.gr

Abstract. The observed relation between prosodic information and the degree of precision and lack of ambiguity is attempted to be integrated in the processing of the user's spoken input in the CitizenShield ("POLIAS") system for consumer complaints for commercial products. The prosodic information contained in the spoken descriptions provided by the consumers is attempted to be preserved with the use of semantically processable markers, classifiable within an Ontological Framework and signaling prosodic prominence in the speakers spoken input. Semantic processability is related to the reusability and/or extensibility of the present system to multilingual applications or even to other types of monolingual applications.

Keywords: Prosodic prominence, Ontology, Selectional Restrictions, Indexical Interpretation for Emphasis, Deixis, Ambiguity resolution, Spatial Expressions.

1 Introduction

In a Human Computer Interaction (HCI) System involving spoken interaction, prosodic information contained in the users spoken input is often lost. In spoken Greek, prosodic information has shown to contribute both to clarity and to ambiguity resolution and, in contrast, semantics and word order are observed to play a secondary role [2]. The relation between prosodic information and the degree of precision and lack of ambiguity is attempted to be integrated in the processing of the user's spoken input in the CitizenShield ("POLIAS") system for consumer complaints for commercial products (National Project: "Processing of Images, Sound and Language", Meter 3.3 of the National Operational Programme "Information Society", which concerns the Research & Technological Development for the Information Society). The preservation of the prosodic information contained in the spoken descriptions provided by the consumers is attempted to be facilitated with the use of semantically processable markers signaling prosodic prominence in the speakers spoken input. Semantic processability is related to the reusability and/or extensibility of the present system to multilingual applications or even to other types of monolingual applications. The spoken input is recognized by the system's Speech

Recognition (ASR) component and is subsequently entered into the templates of the CitizenShield system's automatically generated complaint form.

2 Outline of the CitizenShield Dialog System

The purpose of the CitizenShield dialog system is to handle routine tasks involving food and manufactured products (namely compliants involving quality, product labels, defects and prices), thus allowing the staff of consumer organisations, such as the EKPIZO organisation, to handle more complex cases, such as complaints involving banks and insurance companies. The CitizenShield dialog system involves a hybrid approach to the processing of speaker spoken input in that it involves both keyword recognition and recording of free spoken input. Keyword recognition largely occurs within a yes-no question sequence of a directed dialog (Figure 1). Free spoken input is recorded within a defined period of time, following a question requiring detailed information and/or detailed descriptions (Figure 1). The use of directed dialogs and yes-no questions aims to the highest possible recognition rate of a very broad and varied user group and, additionally, the use of free spoken input processes the detailed information involved in a complex application such as consumer complaints. All spoken input, whether constituting an answer to a yes-no question or constituting an answer to a question triggering a free-input answer, is automatically directed to the respective templates of a complaint form (Figure 2), which are filled in by the spoken utterances, recognized by the system's Automatic Speech Recognition (ASR) component, which is the point of focus in the present paper.

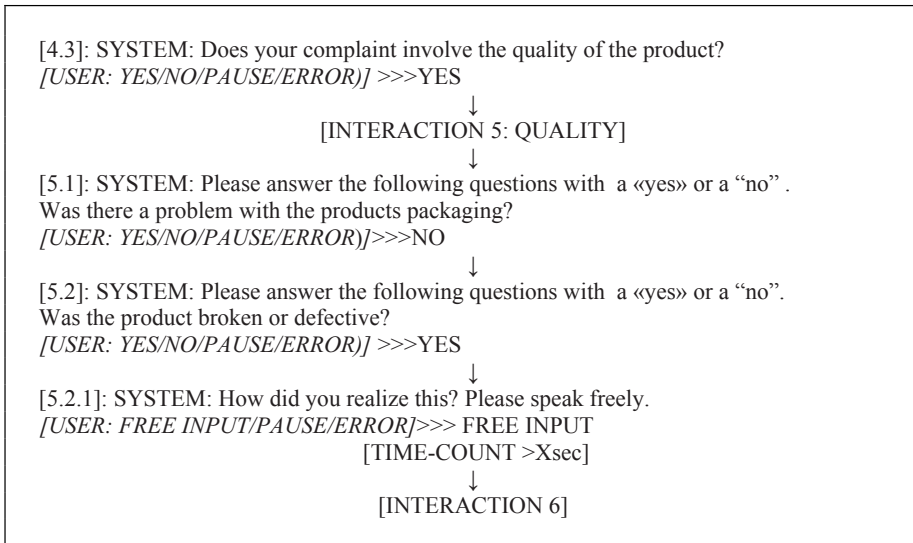


Fig. 1. A section of a directed dialog combining free input (hybrid approach)

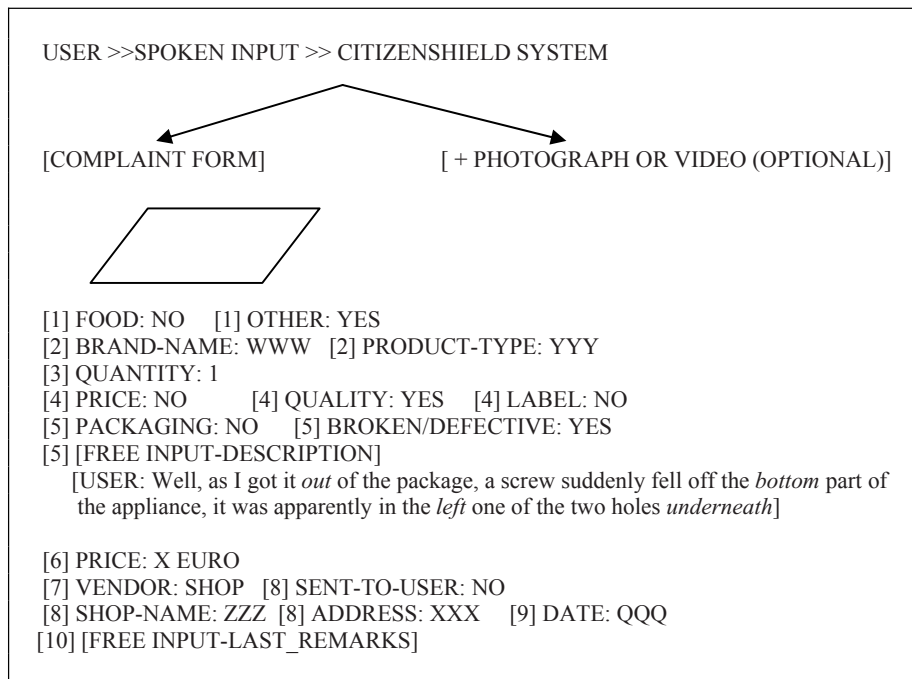


Fig. 2. Example of a section of the data entered in the automatically produced template for consumer complaints in the CitizenShield System (spatial expressions are indicated in italics)

The CitizenShield system offers the user the possibility to provide photographs or videos as an additional input to the system, along with the complaint form. The generation of the template-based complaint forms is also aimed towards the construction of continually updated databases from which statistical and other types of information is retrievable for the use of authorities (for example, the Ministry of Commerce, the Ministry of Health) or other interested parties.

3 Spatial Expressions and Prosodic Prominence

Spatial expressions constitute a common word category encountered in the corpora of user input in the CitizenShield system for consumer complaints, for example, in the description of damages, defects, packaging or product label information. Spatial expressions pose two types of difficulties: They (1) are usually not easily subjected to sublanguage restrictions, in contrast to a significant number of other word-type categories [8], and, (2) Greek spatial expressions, in particular, are often too ambiguous or vague when they are produced outside an in-situ communicative context, where the consumer does not have the possibility to actually “show and tell” his complaints about the product. However, prosodic prominence on the Greek spatial expression has shown to contribute both to the recognition of its “indexical” versus its “vague” interpretation [9], according to previous studies [3], and acts as a default in

preventing its possible interpretation as a part of a quantificational expression—another common word category encountered in the present corpora, since many Greek spatial expressions also occur within a quantificational expression, where usually the quantificational word entity has prosodic prominence.

Specifically, it has been observed that prosodic emphasis or prosodic prominence (1) is equally perceived by most users [2] and (2) contributes to ambiguity resolution of spatial (and temporal) expressions [3].

For the speakers filing consumer complaints, as supported by the working corpus of recorded telephone dialogs (580 dialogs of average length 8 minutes, provided by the speakers belonging to the group of the 1500- 2800 consumers and also registered members of the EKPIZO organization), the use of prosodic prominence helps the user indicate the exact point of the product in which the problem is located, without the help (or, for the future users of the CitizenShield system, with the help) of any accompanying visual material, such as a photograph or a video.

An initial (“start-up”) evaluation of the effect of written texts to be produced by the system’s ASR component where prosodic prominence of spatial expressions is designed to be marked, was performed with a set of sentences expressing descriptions of problematic products and containing the Greek (vague) spatial expressions “on”, “next”, “round” and “in”. For each sentence there was a variant where (a) the spatial expression was signalized in bold print and another variant where (b) the subject or object of the description was signalized in bold print. Thirty (30) subjects, all Greek native speakers (male and female, of average age 29), were asked to write down any spontaneous comments in respect to the given sentences and their variants. 68,3% of the students differentiated a more “exact” interpretation in all (47,3%) or in approximately half (21%) of the sentences where the spatial expressions were signalized in bold print, where 31,5% indicated this differentiation in less than half of the sentences (21%) or in none (10,5%) of the sentences. Of the comments provided, 57,8% focused on a differentiation that may be described as a differentiation between “object of concern” and “point of concern”, while 10,5% expressed discourse-oriented criteria such as “indignation/surprise” versus “description/indication of problem”. We note that in our results we did not take into account the percentage of the subjects (31,5%) that did not provide any comments or very poor feedback.

The indexical interpretation of the spatial expression, related to prosodic prominence (emphasis), may be differentiated in three types of categories, namely (1) indexical interpretation for emphasizing information, (2), indexical interpretation for ambiguity resolution and (3) indexical interpretation for deixis. An example of indexical interpretation for emphasizing information is the prosodic prominence of the spatial expression “mesa” (“in” versus “right in” (with prosodic prominence)) to express that the defective button was sunken right in the interior of the appliance, so that it was, in addition, hard to remove. Examples of indexical interpretation for ambiguity resolution are the spatial expressions “pano” (“on” versus “over” (with prosodic prominence)), “giro” (“round” versus “around” (with prosodic prominence)) and “dipla” (“next-to” versus “along” (with prosodic prominence)) for the respective cases in which the more expensive price was inscribed exactly over the older price elements, the case in which the mould in the spoilt product is detectable exactly at the rim of the jar or container (and not around the container, so it was not easily visible) and the case in which the crack in the coffee machines pot was exactly

parallel to the band in the packaging so it was rendered invisible. Finally, a commonly occurring example of an indexical interpretation for deixis is the spatial expression "e'do"/"e'ki" ("there"/"here" versus right/exactly here/there" (with prosodic prominence)) in the case in which some pictures may not be clear enough and the deictic effect of the emphasized-indexical elements results to the pointing out of the specific problem or detail detected in the picture/video and not to the picture/video in general. With the use of prosodic prominence, the user is able to enhance his or her demonstration of the problem depicted on the photograph or video or describe it in a more efficient way in the (more common case) in which the complaint is not accompanied by any visual material.

The "indexical" interpretation of a spatial expression receiving prosodic prominence can be expressed with the [+ indexical] feature, whereas, the more "vague" interpretation of the same, unemphasized spatial or temporal expression can be expressed with the [- indexical] feature [3]. Thus, in the framework of the CitizenShield system, to account for prosody-dependent indexical versus vague interpretations for Greek spatial expressions, the prosodic prominence of the marked spatial expression is linked to the semantic feature [+ indexical]. If a spatial expression is not prosodically marked, it is linked by default to the [-indexical] feature. In the CitizenShield system's Speech Recognition (ASR) component, prosodically marked words may be in the form of distinctively highlighted words (for instance, bold print or underlined) in the recognized spoken text. Therefore, the recognized text containing the prosodically prominent spatial expression linked to the [+ indexical] feature is entered into the corresponding template of the system's automatic complaint generation form. The text entered in the complaint form is subjected to the necessary manual (or automatic) editing involving the rephrasing of the marked spatial expression to express its indexical interpretation.

In the case of a possible translation of the complaint forms -or even in a multilingual extension of the system, the indexical markers aid the translator to provide the appropriate transfer of the filed complaint, with the respective semantic equivalency and discourse elements, avoiding any possible discrepancies between Greek and any other language.

4 Integrating Prosodic Information Within an Ontological Framework of Spatial Expressions

Since the proposed above-presented prosodic markers are related to the semantic content of the recognized utterance, they may be categorized as semantic entities within an established ontological framework of spatial expressions, also described in the present study. For instance, in the example of the Greek spatial expression "mesa" ("in") the more restrictive concepts can be defined with the features [\pm movement] and [\pm entering area], corresponding to the interpretations "into", "through", "within" and "inside", according to the combination of features used. The features defining each spatial expression, ranging from the more general to the more restrictive spatial concept, are formalized from standard and formal definitions and examples from dictionaries, a methodology encountered in data mining applications [7]. The prosody-dependent indexical versus vague interpretation of these spatial expressions

is accounted for in the form of additional [\pm indexical] features located at the end-nodes of the spatial ontology. Therefore, the semantics are very restricted at the end-nodes of the ontology, accounting for a semantic prominence imitating the prosodic prominence in spoken texts. The level of the [\pm indexical] features may also be regarded as a boundary between the Semantic Level and the Prosodic Level.

Specifically, in the present study, we propose that the semantic information conveyed by prosodic prominence can be established in written texts though the use of modifiers. These modifiers are not randomly used, but constitute an indexical ([+indexical]) interpretation, namely the most restrictive interpretation of the spatial expression in question in respect to the hierarchical framework of an ontology. Thus, the modifiers function as additional semantic restrictions or “Selectional Restrictions” [11], [4] within an ontology of spatial expressions.

Selectional Restrictions, already existing in a less formal manner in the taxonomies of the sciences and in the sublanguages of in non-literary and especially, scientific texts, are applied within an ontology-search tree which provides a hierarchical structure to account for the relation between the concepts with the more general (“vague”) semantic meaning and the concepts with the more restricted (“indexical”) meaning. This mechanism can also account for the relation between spatial expressions with the more general (“vague”) semantic meaning and the spatial expressions with the more restricted (“indexical”) meaning. Additionally, the hierarchical structure, characterizing an ontology, can provide a context-independent framework for describing the sublanguage-independent word category of spatial expressions.

For example, the spatial expression “mesa” (“in”) (Figure 3) can be defined either with the feature (a) [-movement], the feature (b) [+movement] or with the feature (c) [\pm movement]. If the spatial expression involves movement, it can be matched with the English spatial expressions “into”, “through” and “across” [10]. If the spatial expression does not involve movement, it can be matched with the English spatial expressions “within”, “inside” and “indoors” [10]. The corresponding English spatial expressions, in turn, are linked to additional feature structures, as the search is continued further down the ontology. The spatial expression “into” receives the additional feature [+ point] while the spatial expressions “through” and “across”, receive the features [+ area], [\pm horizontal movement] and [+ area], [+ horizontal movement] respectively. The spatial expressions with the [-movement] feature, namely, the expressions, “within”, “inside” and “indoors”, receive the additional feature [+ building] for “indoors”, while the spatial expressions “within” and “inside”, receive the features [\pm object] and [+ object] respectively. The English spatial expression “in” may either signify a specific location and not involve movement, or, in other cases, may involve movement towards a location. All the above-presented spatial expressions can be subject to receive additional restrictions with the feature [+ indexical] as the syntactically realized adverbial modifier “exactly”. It should be noted that the English spatial expressions with an indefinite “ \pm ” value, namely “in”, “through” and “within” also occur as temporal expressions.

To account for prosodically determined indexical versus vague interpretations for the spatial expressions, additional end-nodes with the feature [+ indexical] are added in the respective ontologies, constituting additional Selectional Restrictions. These end-nodes correspond to the terms with the most restrictive semantics to which the

adverbial modifier “exactly” (“akri’vos”) is added to the spatial expression [1]. With this strategy, the modifier “exactly” imitates the prosodic emphasis on the spatial or temporal expression. Therefore, semantic prominence, in the form of Selectional Restrictions located at the end-nodes of the ontology, is linked to prosodic prominence. The semantics are, therefore so restricted at the end-nodes of the ontologies, that they achieve a semantic prominence imitating the prosodic prominence in spoken texts. The adverbial modifier (“exactly”-“akri’vos”) is transformed into a “semantic intensifier”. Within the framework of the rather technical nature of descriptive texts, the modifier-intensifier relation contributes to precision and directness aimed towards the end-user of the text and constitutes a prosody-dependent means of disambiguation.

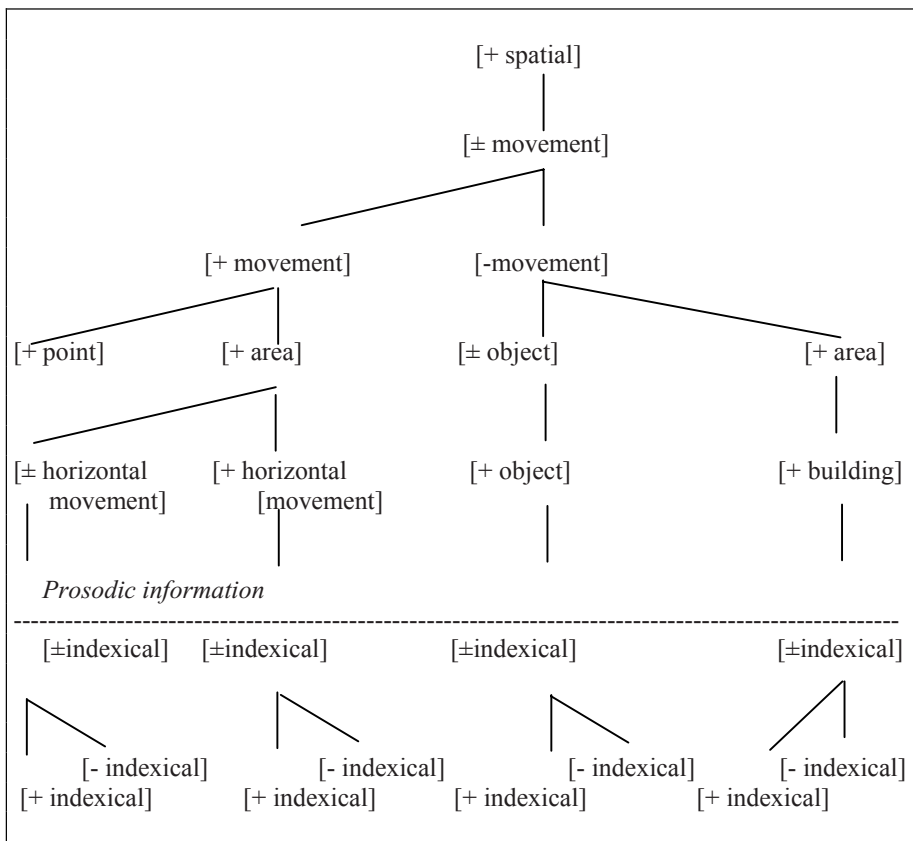


Fig. 3. The Ontology with Selectional Restrictions for the temporal expression “mesa” (“in”)

Therefore, we propose an integration of the use of modifiers acting as Selectional Restrictions for achieving the same effect in written descriptions as it is observed in spoken descriptions, namely directness, clarity, precision and lack of ambiguity.

Specifically, the proposed approach targets to the achievement of the effect of spoken descriptions in a in-situ communicative context with the use of modifiers acting as Selectional Restrictions, located at the end-nodes of the ontologies.

5 Semantically Processable Prosodic Markers Within a Multilingual Extension of the CitizenShield System

The categorization as semantic entities within an ontological framework facilitates the use of the proposed [\pm indexical] features as prosodic markers to be used in the interlinguas of multilingual HCI systems, such as a possible multilingual extension of the CitizenShield system for consumer complaints. An ontological framework will assist in cases where Greek spatial expressions display a larger polysemy and greater ambiguity than in another language (as, for instance, in the language pair English-Greek) and vice versa. Additionally, it is worth noting that when English spatial expressions are used outside the spatial and temporal framework in which they are produced, namely, when they occur in written texts, they, as well, are often too vague or ambiguous. Examples of ambiguities in spatial expressions are the English prepositions classified as Primary Motion Misfits [6]. Examples of “Primary Motion Misfits” are the prepositions “about”, “around”, “over”, “off” and “through”.

Typical examples of the observed relationship between English and Greek spatial expressions are the spatial expressions “dipla”, “mesa”, “giro” with the respective multiple semantic equivalents, namely ‘beside’, ‘at the side of’, ‘nearby’, ‘close by’, ‘next to’ (among others) for the spatial expression “dipla” and ‘in’, ‘into’, ‘inside’, ‘within’ (among others) for the spatial expression “mesa” and, finally, ‘round’, ‘around’, ‘about’ and ‘surrounding’ for the spatial expression “giro” [10]. Another typical example of the broader semantic range of the Greek spatial expressions in respect to English is the term “kato” which, in its strictly locative sense -and not in its quantificational sense, is equivalent to ‘down’, ‘under’, ‘below’ and ‘beneath’.

In a possible multilingual extension of the CitizenShield system producing translated complaint forms (from Greek to another language, for example, English), the answers to yes-no questions may be processed by interlinguas, while the free input (“show and tell”) questions may be subjected to Machine Assisted Translation (MAT) and to possible editing by a human translator, if necessary. Thus, the spatial expressions marked with the [+indexical] feature, related to prosodic emphasis, assist the MAT system and/or the human translator to provide the appropriate rendering of the spatial expression in the target language, whether it used purely for emphasis (1), for ambiguity resolution (2), or for deixis (3). Thus, the above-presented processing of the spatial expressions in the target language contributes to the Information Management during the Translation Process [5]. The translated text, that may accompany photographs or videos, provides detailed information of the consumer’s actual experience. The differences between the phrases containing spatial expressions with prosodic prominence and [+indexical] interpretation and the phrases with the spatial expression without prosodic prominence are described in Figure 4 (prosodic prominence is underlined).

1. Emphasis:

“mesa” = “in”: [“the defective button was sunken in the appliance”]

“mesa” [+indexical] = “right in”:

[“the defective button was sunken right in (the interior) of the appliance”]

2. Ambiguity resolution:

(a) “pano” = “on”: [“the more expensive price was inscribed on the older price”]

“pano” [+ indexical] = “over”:

[“the more expensive price was inscribed exactly over the older price”]

(b) “giro” = “round”: [“the mould was detectable round the rim of the jar”]

“giro” [+ indexical] = “around”:

[“the mould was detectable exactly around the rim of the jar”]

(c) “dipla” = “next-to”: [“the crack was next to the band in the packaging”]

“dipla” [+ indexical] = “along”:

[“the crack was exactly along (parallel) to the band in the packaging”]

3. Deixis:

“e’do”/“e’ki” = “there”/“here” = [“this picture/ video”]

“e’do”/“e’ki” [+ indexical] = “there”/“here” = [“in this picture/video”]

Fig. 4. Marked multiple readings in the recognized text (ASR Component) for translation processing in a Multilingual Extension of the CitizenShield System

6 Conclusions and Further Research

In the proposed approach, the use of semantically processable markers signaling prosodic prominence in the speakers spoken input, recognized by the Automatic Speech Recognition (ASR) component of the system and subsequently entered into an automatically generated complaint form, is aimed to the preservation of the prosodic information contained in the spoken descriptions of problematic products provided by the users. Specifically, the prosodic element of emphasis contributing to directness and precision observed in spatial expressions produced in spoken language are transformed into the [+ indexical] semantic feature. The indexical interpretations of spatial expressions in the present application studied are observed to be differentiated into three categories, namely indexical features used purely for emphasis (1), for ambiguity resolution (2), or for deixis (3). The semantic features are expressed in the form of Selectional Restrictions operating within an ontology. Similar approaches may be examined for other word categories constituting crucial word groups in other spoken text types, and possibly in other languages, in an extended multilingual version of the CitizenShield system.

Acknowledgements. We wish to thank Mr. Ilias Koukoyannis and the Staff of the EKPIZO Consumer Organization for their contribution of crucial importance to the development of the CitizenShield System.

References

1. Alexandris, C.: English as an intervening language in texts of Asian industrial products: Linguistic Strategies in technical translation for less-used European languages. In: Proceedings of the Japanese Society for Language Sciences-JSLS 2005, Tokyo, Japan, pp. 91–94 (2005)
2. Alexandris, C., Fotinea, S-E.: Prosodic Emphasis versus Word Order in Greek Instructive Texts. In: Botinis, A. (ed.): Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics. Athens, Greece, pp. 65–68 (August 28-30, 2006)
3. Alexandris, C., Fotinea, S.-E., Efthimiou, E.: Emphasis as an Extra-Linguistic Marker for Resolving Spatial and Temporal Ambiguities in Machine Translation for a Speech-to-Speech System involving Greek. In: Proceedings of the 3rd International Conference on Universal Access in Human-Computer Interaction (UAHCI 2005), Las Vegas, Nevada, USA (July 22-27, 2005)
4. Gayral, F., Pernelle, N., Saint-Dizier, P.: On Verb Selectional Restrictions: Advantages and Limitations. In: Christodoulakis, D.N. (ed.) NLP 2000. LNCS (LNAI), vol. 1835, pp. 57–68. Springer, Heidelberg (2000)
5. Hatim, B.: Communication Across Cultures: Translation Theory and Contrastive Text Linguistics, University of Exeter Press (1997)
6. Herskovits, A.: Language, Spatial Cognition and Vision, In: Stock, O. (ed.) Spatial and Temporal Reasoning, Kluwer, Boston (1997)
7. Kontos, J., Malagardi, I., Alexandris, C., Bouligaraki, M.: Greek Verb Semantic Processing for Stock Market Text Mining. In: Christodoulakis, D.N. (ed.) NLP 2000. LNCS (LNAI), vol. 1835, pp. 395–405. Springer, Heidelberg (2000)
8. Reuther, U.: Controlling Language in an Industrial Application. In: Proceedings of the Second International Workshop on Controlled Language Applications (CLAW 98), Pittsburgh, pp. 174–183 (1998)
9. Schilder, F., Habel, C.: From Temporal Expressions to Temporal Information: Semantic tagging of News Messages. In: Proceedings of the ACL-2001, Workshop on Temporal and Spatial Information Processing, Pennsylvania, pp. 1309–1316 (2001)
10. Stavropoulos, D.N. (ed.): Oxford Greek-English Learners Dictionary. Oxford (1988)
11. Wilks, Y., Fass, D.: The Preference Semantics Family. In: Computers Math. Applications, vol. 23(2-5), pp. 205–221. Pergamon Press, Amsterdam (1992)