

# Linguistic Processing of Implied Information and Connotative Features in Multilingual HCI Applications

Christina Alexandris and Ioanna Malagardi

National University of Athens, Greece  
calexandris@gs.uoa.gr, i.malagardi@gsrt.gr

**Abstract.** Implied information and connotative features may not always be easily detected or processed in multilingual Human-Computer Interaction Systems for the International Public, especially in applications related to the Service Sector. The proposed filter concerns the detection of implied information and connotative features in HCI applications processing online texts and may be compatible with Interlinguas including the signalization of connotative features, if necessary. The proposed approach combines features detected in the lexical and morpho-syntactic level, and in the prosodic and paralinguistic levels.

**Keywords:** Gricean Cooperativity Principle, online texts, Interlinguas, Morphology, prosodic and paralinguistic features.

## 1 Introduction

The management of implied information and connotative features concerns the Lexical-Semantic - Morphosyntactic Level as well as the Prosodic Level and the Paralinguistic Level. For the management of these features, three types of strategies may be differentiated, depending on the type of application and task-type: excluding, detecting or integrating implied information and connotative features. The exclusion of implied information and connotative features is typically applied in task-related monolingual or multilingual applications in Human Computer Interaction (HCI) Systems, including monolingual Dialog Systems and HCI Systems for the International Public based on Machine Translation (MT). Such applications often involve Controlled Languages and/or a Directed Dialog and System-Initiative-based [12] approach.

Furthermore, some types of monolingual Dialog Systems or other monolingual Human Computer Interaction Systems may pose problems for the International Public, especially in the service sector, as Human-Computer interactions go global [6] and the alternative practice of monolingual variations in the respective languages of a standardized HCI System is proposed. This practice often involves the integration of the language and culture-specific implied information and connotative features.

However, implied information and connotative features may not always be easily avoided in multilingual Human-Computer Interaction Systems for the International Public, especially in applications related to the service sector. In some cases, it may be necessary for such HCI Systems using Machine Translation to include the strategy of detecting implied information and connotative features, according to the type and purpose of the application.

## 2 Application Types

The proposed filter may be compatible with the processing of online texts for the detection of implied information and connotative features in interactive HCI applications processing monolingual online texts for the International Public. This is especially important for professionals, such as journalists, economists and other professionals working with multilingual written and spoken texts, either in specialized domains or in general fields such as business meetings, business transactions and online financial news. It should be noted that International Users may be very fluent in a foreign language or more than one foreign languages, but are either non-native speakers and/or often lack the necessary exposure to the culture related to a foreign language concerned to easily perceive all types of implied information and connotative features.

In online texts, the word groups with implied information and connotative features may be automatically signaled by an online tagger in the proposed filter and in some cases, they be also be provided in combination with the proposed filter as a set of guidelines to International Users, activated according to request.

The proposed filter may also be compatible with Interlinguas. Traditional Interlinguas are geared towards the exclusion and of connotative features [11] [15]. However, in the present approach connotative features may be signaled in Interlinguas, if necessary.

Interlinguas may allow extra tags at a lexical level to be placed with connotative features that may correspond to word groups presented. The connotative features may be automatically signaled operating on detection at morpheme-level or word-level, based on interaction with a database. In addition to detecting implied information and connotative features, the signalization of language-specific prosodic emphasis and other paralinguistic elements at word-level in an Interlingua may play a significant role in the information content of spoken utterances, possibly also in Tonal Languages such as Thai or Chinese.

For all applications concerned, the language-specific “filter” is proposed for the detection or for the integration of implied information and connotative features which may be adapted to the needs of the HCI applications concerned. The filter may be activated or deactivated when processing online texts or during spoken interaction with Interlinguas.

Specifically, in the approach presented, the database and respective tag-set with the word groups and related elements concerned may be used in a variety of multilingual applications, in particular, in the interactive processing of online written and spoken texts, as well as in the processing of Interlinguas and/or its integration in monolingual variations of dialogs. The proposed database and tag-set “filter” is based on features from English, German and Modern Greek, however, it can be adapted and extended to other languages and language groups.

**Table 1.** Interaction with database and tag-set “filter”

<b>Input Type (During Interaction):</b>	<b>Filter:</b>
<b>INTERLINGUAS (ILT OR S-ILT)</b> Speech-to-Speech Translation	<b>Activate – Deactivate</b> ⊗ [open/close]
<b>ONLINE TEXT</b>	

### 3 Filter for Linguistic Processing

The proposed language-specific “filter” constitutes a simple and extendable database constructed on ontological and pragmatic principles [1] [3] and respective tag-set with basic types of word groups related to implied information and connotative features. In particular, the connotative features, implied information and other types of elements may be automatically signalized at morpheme-level or word-level.

The proposed filter is composed of three tiers, the Lexical Tier, the Prosodic Tier and the Paralinguistic Tier and allows the combination of all tiers and respective linguistic and paralinguistic elements to evaluate implied information and connotative features. The combination of all tiers related to the respective Lexical-Semantic - Morphosyntactic Level (Lexical Tier), the Prosodic Level (Prosodic Tier) and the Paralinguistic Level (Paralinguistic Tier) allows a significant extent of coverage of implied information and connotative features for the International Public.

We note that the intensity of the connotative features may be stronger if detected in all three levels, for example, if a word containing (lexical-semantic) connotative features receives prosodic emphasis or is accompanied by additional connotative information in the Paralinguistic Level.

#### 3.1 Tags Related to the Paralinguistic Tier

The Paralinguistic Tier combines paralinguistic elements with spoken words. Paralinguistic Elements are usually language-specific and may vary across cultures. Typical examples of paralinguistic elements are facial expressions such as the raising of eyebrows and frowns, as well as body movements such as gestures related to the hands or nodding of head. Tags for paralinguistic elements may be the annotations “[raising eyebrows]”, “[frown]” and “[nod]”. Additional features may be added, for example, in respect to speed, such as the annotation “[nod-quick]”, in addition to highly language (and culture) specific variations.

#### 3.2 Tags Related to the Prosodic Tier

In the Prosodic Tier, the proposed filter as an annotation module combines spoken words with prosodic elements, such as prosodic emphasis signalizing stressed

elements or a casual attitude: “Stress”, “Casual” [1]. For spoken Machine Translation, types of Interlinguas (ILTs) allowing the recognition and isolated processing of keywords at a lexical level, such as Simple Interlinguas [13], facilitate the signalization of prosodic features detected at word level, such as prosodic emphasis. For example, the marker prosodic emphasis [+STRESS] can be used as an additional paralinguistic marker on the lexical element of the Interlingua or in the online spoken text. The marker [+CASUAL] may be used for a casual tone, whereas the [+NON-NEUTRAL] is used on other types of prosodic elements related to implied connotative features on the Prosodic Level.

### 3.3 Tags Related to the Lexical Tier and Pragmatic Principles

In the Lexical Tier, corresponding to tags concerning the Lexical Level, the word groups with connotative features are differentiated according to criteria related to Pragmatics, namely the flouting of the Maxims stated in the Gricean Cooperativity Principle [8] [9].

In respect to the Gricean Maxim of Quality, namely “Do not say what you believe to be false” and “Do not say that for which you lack adequate evidence”, in the case in which the Maxim is not flouted for the purpose of propaganda, it is observed that information presented in a form characterized as flouting the Gricean Maxim of Quality often contains superfluous elements flouting the Gricean Maxim of Quantity. This relationship applies in a similar way to information flouting the Gricean Maxim of Manner and, specifically, “Avoid obscurity of expression”, “Avoid ambiguity” and “Be orderly”, where, unless propaganda or similar communicative targets are involved, superfluous information (flouting the Gricean Maxim of Quantity) is often encountered in written and spoken texts where there is flouting of the Gricean Maxim of Manner. The Gricean Maxim of Manner also includes the part “Be brief” (avoid unnecessary prolixity) which partially coincides with the Gricean Maxim of Quantity and is thus directly related to the avoidance of superfluous information.

Specifically, these word groups to be detected or integrated in multilingual applications concern both specific types of semantic features related to the superfluous information connected to the above-described Maxims, such as mode, malignant/benign action or emotional/ethical gravity, as well as particular types of grammatical features in verbs, adjectives, adverbials and in specific types of suffixes and particles. The language-specific tag set of word groups ranges from the more evident yet less frequent strong or emotional expressions to the less obvious and commonly-occurring word categories constituting word groups related to implied information and connotative features.

These word categories function as subtle hints or tell-tale signs of connotative elements and implied information and may sometimes be especially problematic to the International Public.

In other words, the overall context of the written and spoken text may be described as containing a subset of word-types, coinciding with superfluous information in the text and indicating emotionally and socio-culturally “marked” elements constituting implied information and connotative features and expressing style and overall spirit of

the author, speaker and the intended readership or audience. Thus, the criteria for determining tagged word types with implied and connotative features are related to Pragmatics, in particular, the flouting of the Gricean Cooperativity Principle [8] [9].

**Table 2.** Tiers of the proposed database and tag-set “filter”

---

**Features:**

**Paralinguistic Tier**  
*(Paralinguistic Level)*  
 [raising eyebrows] [frown] [nod-quick]

**Prosodic Tier**  
*(Prosodic Level)*  
 [±STRESS] [±CASUAL] [±NON-NEUTRAL]  
 (other types of prosodic elements)

**Lexical Tier**  
*(Lexical Level/Word Level – Morphological Level)*  
 [sem-expl-conn]  
 [sem-impl-conn]  
 [prag-conn:±emph]  
 [prag-conn:subj]  
 [prag-conn:modl]

---

## 4 Connotative Feature Types

Tags corresponding to word groups with implied information or connotative features in their semantic content may be divided into word categories where connotative features are detected at a word level (Lexical Tier) and word categories where connotative features are detected in a morphological level (morpheme level) (Lexical Tier). Connotative features detected at word level or at the morphological level are either related to word groups whose semantic content is related to connotatively emotionally, and socio-culturally “marked” elements (Semantic Content categories) or word groups whose pragmatic usage concerns connotative features and implied information (Pragmatic Usage categories). From the aspect of Prosody, it is observed that the word categories detected at a word level are sensitive to prosodic emphasis. In contrary, the word categories detected at the morphological level are not affected by prosodic emphasis.

### 4.1 Word Level and Semantic Content

Word categories with connotative features detected at the word level (Lexical Tier) and whose semantic content is related to connotatively emotionally, and socio-culturally “marked” elements constitute word groups with evident or explicit connotative features.

Typical examples of word groups with explicit connotative features are the grammatical categories of adjectives and adverbials, containing semantic features related to (i) descriptive features (ii) mode (iii) malignant/benign action or (iv) emotional/ethical gravity [3].

The evident connotative features of the above-described word categories may be formalized as special features linked to the respective word category type. Specifically, the feature [sem-expl-conn] may be appended to these categories and matched on a word-level, being directly matched to the entire word. In respect to the Prosodic Level of this group of word categories, prosodic emphasis may emphasize or intensify the semantics of the emphasized word without determining the semantic content. The word groups with evident connotative features may be classified as “Prosodically Sensitive” words.

**Table 3.** Connotative Feature Types

		CONNOTATIVE	
		DETECTION:	FEATURE TYPE: PROSODY:
WORD CATEGORY	→	WORD LEVEL →	SEMANTIC CONTENT “Prosodically Sensitive”
		→	PRAGMATIC USAGE “Prosodically Sensitive”
	→	MORPHO- LOGICAL LEVEL →	SEMANTIC CONTENT “Prosodically Independent”
		→	PRAGMATIC USAGE “Prosodically Independent”

### 4.2 Morphological Level and Semantic Content

Word categories whose connotative features are detected in the morphological level (Lexical Tier) and whose semantic content is related to connotatively emotionally and socio-culturally “marked” elements also may be referred to as word groups with implicit connotative features [2].

Word groups with implied connotative features include the grammatical categories of verb-stems (or nominalizations of verbs) containing semantic features (including implied connotations in language use) related to (i) mode (ii) malignant/benign action or (iii) emotional/ethical gravity, as well as nouns with suffixes producing diminutives, derivational suffixes resulting to a (ii) verbalization, (iii) an adjectivization or (iii) an additional nominalization of proper nouns [1] [3].

The implicit connotative features of the above-described word categories may be formalized as special features linked to the respective word category type. Specifically, the feature [sem-impl-conn] may be appended to these categories, matched on a morphological level, on the verb-stem of verbs containing semantic features related to mode, malignant/benign action or emotional/ethical gravity, and on the suffix of nouns with suffixes producing diminutives [2]. Additionally, for the appending of the [sem-impl-conn] feature, verb-stems are compared with derivational suffixes resulting to a nominalization of verbs (excluding derivational suffixes producing participles and actor thematic roles). Stems of proper nouns are compared with derivational suffixes resulting to a verbalization or adjectivization of proper nouns [3].

In respect to the Prosodic Level, the presence or absence of prosodic emphasis on words of this word group 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 this category, which may be classified as “Prosodically Independent”, is sentence dependent and highly sublanguage- and application-specific.

### 4.3 Word Level and Pragmatic Usage: Adverbials and Particles

Word categories with connotative features detected at word level (Lexical Tier) and whose pragmatic use concerns connotative features and implied information involve language-specific sets of adverbials, discourse particles or other language-specific grammatical categories.

The feature [prag-conn:±emph] is matched on a word-level being directly matched to the adverbial or particle used in languages such as English (“so”) or German (“eben”, “gleich”) for an emphatic or casual/spontaneous effect or the discourse particle identified as a “politeness marker” in Modern Greek (“Πείτε μου”).

In spoken language, these adverbials and particles may either be used to emphasize the semantic content of the spoken phrase or sentence (emphasis [+emph]), to allow a more casual or spontaneous effect of the overall spoken utterance (casual, [-emph]) or to achieve politeness (politeness-markers, [-emph])).

Regarding the Prosodic Level, for discourse particles identified as “politeness markers”, the absence of prosodic emphasis signals them as politeness markers, while with the presence of prosodic emphasis they only have the property of discourse particles [4][5]. Similarly, for adverbials and particles in languages such as English

(“so”) or German (“gleich”), the absence of prosodic emphasis signalizes a casual or spontaneous effect, which is not achieved with the presence of prosodic emphasis [3].

Previous studies have demonstrated a differentiation between specific word categories in which prosodic emphasis does not determine their semantic content and word categories whose semantic content may be determined by prosodic emphasis [5]. In the present case involving adverbials and particles, the semantic content is not entirely determined by the presence or absence of prosodic emphasis, however, the pragmatic features within the utterance and the related connotative aspects are affected. The group of word categories whose semantic content may be affected by prosodic emphasis is classified as “Prosodically Sensitive” words [5].

#### 4.4 Morphological Level and Pragmatic Usage: Other Grammatical Features

Word categories with connotative elements detected at the morphological level (Lexical Tier) and whose pragmatic use concerns connotative elements and implied information concern various types of grammatical features. Grammatical features inherently present in languages may contain implied semantic and connotative information which is not always easily detected or successfully managed in the translation process.

Examples of ambiguity related to implied or hypothetical actions are modal verbs and verbal adjectives. Apart from their literal meaning to express a suggestion or a prediction, modal verbs in English and German such as “should” (“soll” in German) or “would” (“wuerde” in German) are often used as understatements, an implied intention, sometimes even irony. These grammatical categories may be in many cases especially problematic to the International Public, both in written language and in spoken language.

Another example of inherent grammatical features and implied information and connotative features are the suffixes in specific verb groups of pro-drop languages. The connotative feature of politeness or friendliness can be expressed in the form of a relationship between subject and object is detected in Greek verb suffixes.

Specifically, we note that, in Greek, as a verb-framed and pro-drop language (like Spanish or Italian), the verbal features in verb’s suffix imply the subject. This morphological characteristic affects the semantics and connotative features of certain verb categories (verb stems), especially verbs expressing a service or any benign action concerning an object or the verb’s subject or both the subject and object of the action expressed. With this way, a relation between the subject and the object is expressed, signaling politeness, especially in spoken language, if receiving prosodic stress [13]. Emphasis is placed on the User’s wish or response. For example, the Greek verbs “theleis” (“[you] want”) and “olokli’rosate” (“[you] have finished-completed” [your input]) is equivalent to the verbs “want” and “finished” in English respectively.

Features appended to these categories, are the feature [prag-conn:subj] is matched on a Morphological Level, being matched on the suffix of verbs of verb-framed and pro-drop languages.

For modal verbs, the subset of modal verbs containing likely connotative features is signalized by the feature [prag-conn:modl] [3]. Feature types allow an automatic



grouping of verb groups and other word groups [10] [14] and may also be retrieved with the help of Wordnets. Additional feature types may be added, according to the language concerned.

In respect to the Prosodic Level, words of this word group are unaffected by the presence or absence of prosodic emphasis in respect to their semantic content, constituting a “Prosodically Independent” category.

As a final comment in respect to the observed “Prosodically Independent” categories related to Semantic Content and to Pragmatic Usage, the Morphological Level appears to be opaque to any prosodic interference affecting semantic content. However, further research is required to evaluate this observation.

## 5 Conclusions and Further Research

The proposed approach involves a database constructed on ontological and pragmatic principles and respective tag-set with basic types of word groups related to implied information and connotative features, constituting a language-specific “filter” for HCI applications.

The proposed language-specific “filter” may operate simultaneously on the Morphosyntactic and Lexical-Semantic Level, the Prosodic Level and the Paralinguistic Level. Additionally, the proposed approach allows the processing of implied information and connotative features related in the combination of multiple linguistic levels, enabling access to complex types of implied information and connotative features.

Furthermore, with the proposed approach is targeted to allow flexibility in respect to various languages, serving as an onset for a cross-linguistic approach. In particular, the language-specific annotation containing implied information and connotative features may be inserted and signaled at the Lexical Tier and the Prosodic Tier, as well as the Paralinguistic Tier, providing a flexible framework for the processing of various languages other than the languages presented.

It additionally may be noted that the management of implied information and connotative features in word groups may contribute to ambiguity resolution in semantic webs and in some cases even in the social semantic web, especially in respect to tags [7].

An additional target is to gain insight for the formalization of a basic framework for processing similar or contrasting linguistic and cultural features of other language families. Further research including a comparison with other languages and language families may allow the integration of additional features and/or aspects in the proposed general framework.

## References

1. Alexandris, C.: User Interface Design for the Interactive Use of Online Spoken German Journalistic Texts for the International Public. In: Stephanidis, C. (ed.) Posters, Part I, HCI 2011. CCIS, vol. 173, pp. 551–555. Springer, Heidelberg (2011)
2. Alexandris, C.: English, German and the International “Semi-professional” Translator: A Morphological Approach to Implied Connotative Features. *Journal of Language and Translation* 11(2), 7–46 (2010)

3. Alexandris, C.: *Speech Acts and Prosodic Modeling in Service-Oriented Dialog Systems*. In: *Computer Science Research and Technology*. Nova Science Publishers, Hauppauge (2010)
4. 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)
5. 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 (2008)
6. Altman-Klein, H., Lippa, K., Lin, M.H.: *As Human-Computer Interactions Go Global*. In: Hayes, C.C., Miller, C.A. (eds.) *Human-Computer Etiquette*, Taylor and Francis, Boca Raton, FL (2011)
7. Breslin, J.G., Passant, A., Decker, S.: *The Social Semantic Web*. Springer, New York (2009)
8. Grice, H.P.: *Logic and conversation*. In: Cole, P., Morgan, J. (eds.) *Syntax and Semantics*, vol. 3. Academic Press, New York (1975)
9. Hatim, B.: *Communication Across Cultures: Translation Theory and Contrastive Text Linguistics*. University of Exeter, Exeter (1997)
10. 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)
11. 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)
12. Jurafsky, D., Martin, J.: *Speech and Language Processing, an Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*, 2nd edn. Prentice Hall series in Artificial Intelligence. Pearson Education, Upper Saddle River (2008)
13. 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)
14. 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>
15. 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)