



Generating Graphic Representations of Spoken Interactions Revisited: The Tension Factor and Information Not Uttered in Journalistic Data

Christina Alexandris^{1,2}, Dimitrios Mourouzidis^{1,2}(✉),
and Vasilios Floros^{1,2}

¹ National and Kapodistrian University of Athens, Athens, Greece
calexandris@gs.uoa.gr, mourouzidisd@gmail.com,
florosbas2002@yahoo.gr

² European Communication Institute (ECI), Danube University Krems
and National Technical University of Athens, Athens, Greece

Abstract. The proposed interactive and semi-automatic processing in distinctive modules facilitates the correct perception and evaluation of pragmatic features and paralinguistic features in spoken interaction, especially in discussions and interactions beyond a defined agenda and specified protocol, such as interviews and live conversations in Skype or in the Media. We propose a processing and evaluation framework including a generation of graphical representations and tags corresponding to values and benchmarks depicting the degree of information not uttered and non-neutral elements - including tension - in Speaker behavior in spoken text segments.

Keywords: Spoken journalistic texts · Spoken interaction · Tension · Paralinguistic features · Graphic representations · Cognitive Bias

1 Processing Information Not Uttered in Spoken Journalistic Texts

Pragmatic features in spoken interaction and information conveyed but not uttered by Speakers can pose challenges to applications processing spoken texts that are not domain-specific. The proposed interactive and semi-automatic processing in distinctive modules facilitates the correct perception and evaluation of pragmatic features and paralinguistic features in spoken interaction, especially in discussions and interactions beyond a defined agenda and specified protocol, such as interviews and live conversations in Skype or in the Media.

We propose a processing and evaluation framework including a generation of graphical representations and tags corresponding to values and benchmarks depicting the degree of information not uttered and non-neutral elements in Speaker behavior in spoken text segments. Special focus is placed on the element of tension. The generated tags and values can be used for text classification for the development and collection of empirical data for HCI and HRI applications and for applications such as Sentiment Analysis and Opinion Mining.

Spoken political and journalistic texts may be considered to be a remarkable source of empirical data both for human behaviour and for linguistic phenomena, especially for spoken language. However, with some exceptions, spoken political and journalistic texts are usually underrepresented both in linguistic data for translational and analysis purposes and in Natural Language Processing (NLP) applications. These text types pose challenges for their evaluation, processing and translation since they are usually rich in socio-linguistic and socio-cultural elements, include discussions and interactions beyond a defined agenda and are often not domain-specific. Furthermore, with spoken political and journalistic texts there is always the possibility of different types of targeted audiences - including non-native speakers and the international community. In these cases, essential information, presented either in a subtle form or in an indirect way, is often undetected, especially by the international public.

As the variety and complexity of spoken Human Computer Interaction (HCI) (and Human Robot Interaction - HRI) applications increases, the correct perception and evaluation of information not uttered is an essential requirement in systems with emotion recognition, virtual negotiation, psychological support or decision-making.

Furthermore, Information that is not uttered is problematic in Data Mining and Opinion Mining applications, since they mostly rely on word groups, word sequences and/or sentiment lexica [18], including recent approaches with the use of neural networks [6, 15, 29]. In recent research for Sentiment Analysis from videos (text, audio and video) with the use of a hierarchical architecture for extracting context dependent multimodal utterance features [26], it was observed that, in some cases, the gesture, facial expression or movement may either complement or contradict the semantic content of a spoken utterance, even in domain-specific applications.

The graphic patterns and visual representations are based on the output of an interactive annotation tool for spoken journalistic texts presented in previous research [4]. Specifically, in the interactive annotation tool [4], incoming texts to be processed constitute transcribed data from journalistic texts. The annotation tool was designed to operate with most commercial transcription tools, some of which are available online. The development of the tool is based on data and observations provided by professional journalists (European Communication Institute, Program M.A in Quality Journalism and Digital Technologies, Danube University at Krems, Athena- Research and Innovation Center in Information, Communication and Knowledge Technologies, Athens - Institution of Promotion of Journalism Ath.Vas. Botsi, Athens and the National and Technical University of Athens, Greece). Since processing speed and the option of re-usability in multiple languages of the written and spoken political and journalistic texts constitutes a basic target of the proposed approach, strategies typically employed in the construction of Spoken Dialog Systems, such as keyword processing in the form of topic detection, were adapted in the developed annotation tool. The functions of the designed and constructed interactive annotation tool [4] include providing the User-Journalist with (a) the tracked indications of the topics handled in the interview or discussion and (b) the graphic pattern of the discourse structure of the interview or discussion. Furthermore, these functions facilitate the comparison between discourse structures of conversations and interviews with similar topics or the same participants/participant.

2 Generated Graphical Representations and Tags: The “Relevance” Module and Previous Research

Generated graphical representations and annotation options are proposed for identifying the complex types of information presented, in combination to the respective activated modules within a singular annotation and processing framework. All strategies and respective modules presented are based on the Gricean Cooperative Principle [12, 13] in the Speech Acts involved.

Pragmatic features, in particular, indicators of a Speaker’s attitude-behavior and intentions, including tension, can be visualized in distinctive generated graphic representations and related annotations. The generated distinct types of graphic patterns presented here contribute to a user-independent evaluation of spoken Human-Human conversation and interaction [3, 21].

In small speech segments with constant and quick change of speaker turns and with discourse structure compatible to models where each participant selects self [27, 34], topic tracking (and topic change) allows the evaluation of speaker behavior and enables the identification of Speaker’s intentions and Illocutionary Speech Acts performed [7, 28]. Topic tracking can be applied especially in short speech segments with two or multiple Speakers-Participants [3]. The content of relatively short utterances can be summarized with the use of keywords chosen from each utterance by the user-evaluator [3], with the assistance of the Stanford POS Tagger for the automatic signalization of nouns in each turn taken by the Speakers in the respective segment in the dialog structure. The registered and tracked keywords, treated as local variables, signalize each topic and the relations between topics, since automatic Rhetorical Structure Theory (RST) analysis procedures [30, 36] usually involves larger (written) texts and may not produce the required results.

The implemented “RELEVANCE” Module [21] generates a visual representation from the user’s interaction, tracking the corresponding selected topic-keywords in the dialog flow, as well as the chosen types of relations between them. The interactive generation of registered paths is similar to the paths with generated sequences of recognized keywords in spoken dialog systems, in the domains of consumer complaints and mobile phone services call centers [11, 23]. This function is similar to user-independent evaluations of spoken dialog systems [33] for by-passing User bias [9, 22]. Keywords (topics) may be repeated or related to a more general concept (or global variable) [17] or related to keywords (topics) concerning similar functions (corresponding to the Repetition, Generalization and Association relations respectively and the visual representations of Distances 1 (value “1”), 2 (value “2”) and 3 (value “3”) respectively) [3]. A keyword involving a new command or function is registered as a new topic (New Topic, visual representation of Distance 4, corresponding to value: “0”). The sequence of topics chosen by the user and the perceived relations between them generates a “path” of interaction, forming distinctive visual representations stored in a database currently under development: Topics and words generating diverse reactions and choices from users result to the generation of different forms of generated visual representations for the same conversation and interaction [3, 21].

The generated visual representations depict topics avoided, introduced or repeatedly referred to by each Speaker-Participant, and in specific types of cases may indicate the existence of additional, “hidden” Illocutionary Acts other than “Obtaining Information Asked” or “Providing Information Asked” in a discussion or interview. Thus, the evaluation of Speaker-Participant behavior targets to by-pass Cognitive Bias, specifically, Confidence Bias [16] of the user-evaluator, especially if multiple users-evaluators may produce different forms of generated visual representations for the same conversation and interaction and compared to each other in the database. In this case, chosen relations between topics may describe Lexical Bias [31] and may differ according to political, socio-cultural and linguistic characteristics of the user-evaluator, especially if international users are concerned [5, 19, 25, 35] due to lack of world knowledge of the language community involved [14, 24, 32]. The envisioned further development of generated visual representations is their modeling in a form of graphs, similar to discourse trees [8, 20].

The types of relations-distances between word-topics chosen by the user-evaluator are registered and counted. If the number of (a) the “Repetitions” or (b) the number of the “Generalizations” or (c) the number of the “Topic Switches” exceeds well over 50% of the registered relations-distances between word-topics, the interaction is signalized for further evaluation, containing Illocutionary Acts not restricted to “Obtaining Information Asked” or “Providing Information Asked”. The following benchmarks indicate interactions with Illocutionary Acts beyond the predefined framework of the dialog for multiple Speaker discussions and/or short speech segments, where D_s = Number of Distances and S_p = Number of Speaker turns [1]:

- $X = D_s \leq S_p$ (calculating over 50% of “Repetitions” (Distance = 1, value “1”) or “Topic Switches” (Distance = 4, value “0”).
- $X = D_s > S_p \times Gen$ ($Gen = S_p \times 3 \div 2$) (calculating over 50% of “Generalizations” (Distance = 3, value “3”).

These benchmarks for dialogs with short speech segments can be referred to as “(Topic) Relevance” benchmarks with a value of “X” or “Relevance (X)” [1].

The above-described values, benchmarks [1] and graphic representations also allow the identification and detection of additional, “hidden” Illocutionary Acts not restricted to “Obtaining Information Asked” or “Providing Information Asked”, as defined by the framework of the interview or discussion [21]. Three frequently detected categories of pointers to “hidden” Speech Acts are: “Presence” (reluctance to answer questions, avoidance of topics, polite or symbolic presence in the discussion or interview but not an active participation), “Express Policy” (direct or even blatant expression of opinion or policy- persistence on discussing the same topic of interest or attempts to direct the discussion in the topic(s) or interest) and “Make Impression” (behavior similar to the previous categories - with characteristic prosodic and paralinguistic features). These Speech Act pointers may be connected to each other and may even occur at the same time. The “Make Impression” Speech Act pointer is distinguished from the other two Speech Act pointer since it is identifiable on the Paralinguistic Level [21].

The “[IMPL]” tag is generated after the activation of the above-described “RELEVANCE” Module signaling the presence of additional, “hidden” Illocutionary Acts performed by the Speakers-Participants. The “[IMPL]” tag may be accompanied by an indication of the “Presence”, “Express Policy” or “Make Impression” Speech Act pointer, if applicable. Figure 1 and 2 depict graphical representations of the “RELEVANCE” Module Output: Generated graphical representation with multiple “Topic Switch” relations [21] and generated graphical representation with multiple “Generalization” relations [21], both resulting to the generation of the “[IMPL]” tag.

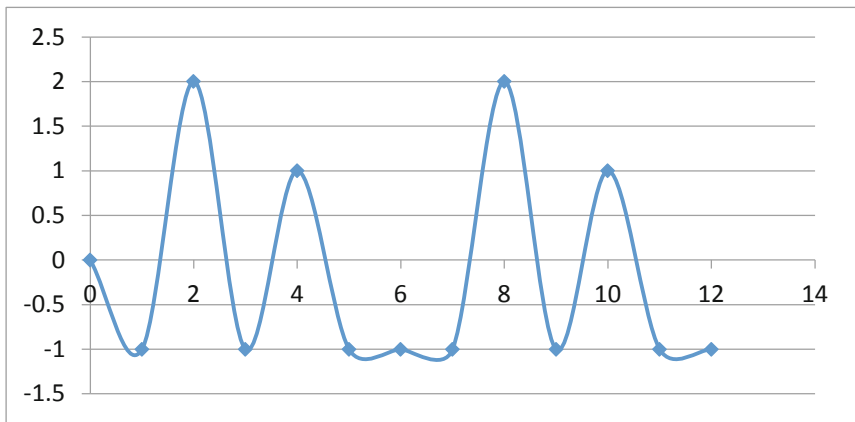


Fig. 1. Generated graphical representation with multiple “Topic Switch” relations (Mourouzidis et al., 2019) producing the [IMPL] tag as output.

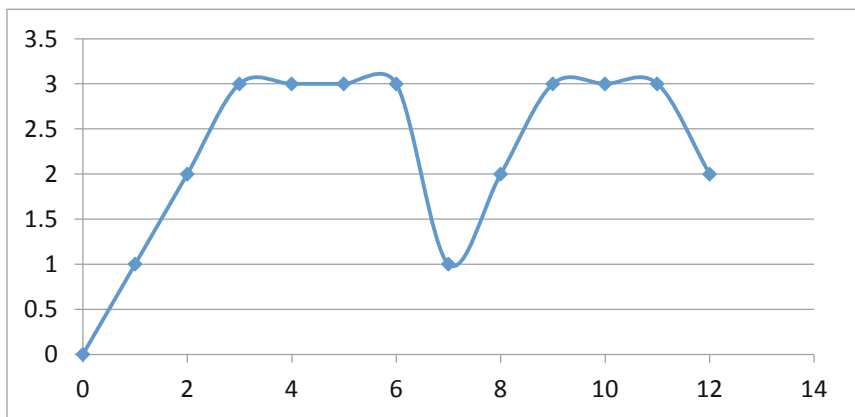


Fig. 2. Generated graphical representation with multiple “Generalization” relations (Mourouzidis et al., 2019) producing the [IMPL] tag as output.

3 Generating Graphical Representations Revisited: The Tension Factor

The further development of the database containing registered spoken interaction for determining and evaluating Cognitive Bias in spoken journalistic texts [3, 21] involves the processing of discussions and interviews containing larger speech segments. In the case of discussions and interviews containing larger speech segments, the identification of speaker's intentions and "hidden" Illocutionary Act detection follows a process locating points of possible tension and/or conflict between speakers-participants. In points of possible tension and/or conflict between speakers-participants, Cognitive Bias can both be by-passed or registered. Cognitive Bias is by-passed by signaling and counting the points of possible tension and/or conflict between speakers-participants henceforth referred to as "hot spots" [1]. The signalization of "hot spots" is based on the violation of the Quantity, Quality and Manner Maxims of the Gricean Cooperativity Principle [12, 13]. Cognitive Bias is registered by comparing content of the Speaker turns in the signalized "hot spots" and assigning a respective value.

The above-described "Presence" Pointer, in some cases, the "Make Impression" Pointer or "Express Policy" Pointer to the Speaker's intentions and behavior is related to the values of the "Relevance (X)", "Tension (Y)" and "Collaboration (Z)" benchmarks [1]. These benchmarks and related visual representations are based on the Gricean Cooperative Principle and may be used for evaluating the Cognitive Bias-Confidence Bias [16] of the user-evaluator of the recorded and transcribed discussion or interview. Graphic representations and values enable the evaluation of the behavior of speakers-participants, depicting Cognitive Bias and may also serve for by-passing Confidence Bias of the user-evaluator of the recorded and transcribed discussion or interview.

Targeting to by-pass Cognitive Bias in two-party discussions and interviews containing longer speech segments, a proposed semi-automatic procedure, the "TENSION" Module, involves "taking the temperature" of a transcribed dialog by measuring the number of detected points of possible tension and/or conflict between Speakers-Participants, referred to as "hot spots". The signalization of multiple "hot spots" indicates a more argumentative than a collaborative interaction, even if Speakers-Participants display a calm and composed behavior. In particular, the Illocutionary Act performed by the Speaker concerned may not be restricted to "Obtaining Information Asked" or "Providing Information Asked" in a discussion or interview.

A "hot spot" consists of the pair of utterances of both speakers, namely a question-answer pair or a statement-response pair or any other type of relation between speaker turns. In longer utterances, the first 60 words of the second speaker's (Speaker 2) utterance are processed (approximately 1–3 sentences, depending on length, with the average sentence length of 15–20 words, [10] and the last 60 words of the first speaker's (Speaker 1) utterance are processed (approximately 1–3 sentences, depending on length). The automatically signalized "hot spots" are extracted to a separate template for further processing. The extraction contains not only the detected segments but also the complete utterances consisting of both speaker turns of Speaker 1 and Speaker 2. For a segment of speaker turns to be automatically identified as a "hot spot", at least

two of the following three conditions (1), (2) and (3) must apply [1] to one or to both of the speaker's utterances, of which conditions (1), (2) are directly or indirectly related to flouting of Maxims of the Gricean Cooperative Principle [12, 13]. These conditions are the following, with features detectable with a POS Tagger (for example, the Stanford POS Tagger, <http://nlp.stanford.edu/software/tagger.shtml>) or they may constitute a small set of entries in a specially created lexicon or may be retrieved from existing databases or WordNets:

- (1) Additional, modifying features. In one or in both speakers' utterances in the segment of speaker turns there is at least one phrase containing a sequence of two adjectives (ADJ ADJ) (a) or an adverb and an adjective (or more adjectives) (b) (ADV ADJ) or two adverbs (ADV ADV) (c) (Violation of the Gricean Cooperative Principle in respect to the Maxim of Quantity -"Do not make your contribution more informative than is required") [1].
- (2) Reference to the interaction itself and to its participants with negation. For example, "I" or "you" ((I/You) "don't", "do not", "cannot") (a) and in the verb phrase (VP) there is at least one speech-related or behavior verb-stem referring to the dialog itself (b) (for example, "speak", "listen", "guess", "understand") (including to parts of speech other than verbs (i.e. "guessing", "listener"), as well as to words constituting parts of expressions related to speech or behavior ("conclusions", "words", "mouth", "polite", "nonsense", "manners"), (violation of the Gricean Cooperative Principle in respect to the Maxim of Quality -"1- Do not say what you believe to be false", "2 - Do not say that for which you lack adequate evidence") [12, 13] and/or in respect to the Maxim of Manner -Submaxim 2 "Avoid ambiguity") [12, 13] in the utterance of the previous Speaker: considered unacceptable, ambiguous, false or controversial) [1].
- (3) Prosodic emphasis and/or Exclamations. (a) Exclamations include expressions such as "Look", "Wait" and "Stop". (b) Prosodic emphasis, detected in the speech processing module, may occur in one or more of the above-described words of categories (1a, 1b, 1c, 2a and 2b) or in the noun or verb following (modified by) 1a, 1b and 1c [1].

The benchmark for evaluating a remarkable degree of tension in a discussion is signaled by multiple "hot spots" detected and not sporadic occurrences of "hot spots". Thus, the number of 1-2 "hot spot" occurrences in longer speech segments in question (30-45 min) signalizes a low degree of tension. A remarkable degree of tension in a 30-45 min discussion or interview is related to a number of at least 4 detected "hot spots" (where the number of 3 hot spots constitutes a marginal value). Detected points of possible tension and/or conflict are indicated by the following benchmark (where $Y = \text{wav file length in minutes} \div \text{the number of "hot spot" signalized speech segments}$): $Y < 10$. (Example: File length = 35 min, SPEECH SEGMENT-count: 5, Evaluation: 7). These benchmarks for dialogs with long speech segments can be referred to as "Tension" benchmarks with a value of "Y" or "Tension (Y)" [1].

Additionally, each “hot spot” is marked with a (1,1) if both speakers’ utterances are considered equally non-collaborative (1, 0) for Speaker 1 (in this case, the journalist-reporter), (0, 1), if the interviewee’s (Speaker 2) reaction is not justified in respect to the style and content of the utterance of Speaker 1 and (0, 0), if a “hot spot” speech segment is evaluated by the user not as a point of possible tension and/or conflict between speakers-participants (false “hot spot”- [1]).

Both Speakers may have an equal number of a grading of “1” in all extracted “hot spots” detected or one of the Speakers may have a slightly higher/lower or a considerably higher/lower grading of “1”. A grading of “1” in 50% or more of the “hot spots” signalizes that the Illocutionary Act performed by the Speaker concerned is not restricted to “Obtaining Information Asked” or “Providing Information Asked”. Speaker behavior indicating that Illocutionary Acts performed are not restricted to the predefined interaction framework is evaluated by the following benchmarks (where Z = the number of “hot spot” signalized speech segments divided by (÷): 2 (50%): Sum of Speaker grades \geq Z. (Example: SPEAKER1 (1, 1, 1, 0, 1), SPEAKER2 (0, 0, 1, 1, 0), SPEECH-SEGMENT-count “hot spots”: 5, sum of grades = 6, $6 \geq Z$ where $Z = 2.5$). These benchmarks for dialogs with long speech segments can be referred to as “Collaboration” benchmarks with a value of “Z” or “Collaboration (Z)”.

In the proposed annotation options, the [IMPL] tag for text segments at sentence, passage or text level signalizes the presence of “hot spot” as a feature related to complex information content, including implied information, intentions, attitude and behavior.

The “[IMPL]” tag is generated after the activation of the above-described “TENSION” Module (Fig. 3) signalizing a remarkable degree of tension and uncollaborative behavior between the Speakers-Participants and the presence of additional, “hidden” Illocutionary Acts performed (Figs. 4 and 5).

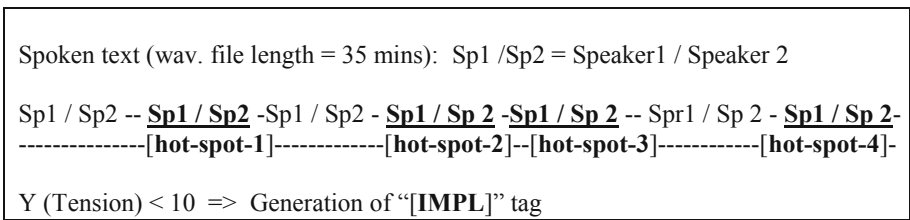


Fig. 3. “TENSION” Module Output: Signalization of multiple “hot spots” in a spoken text segment for the generation of the “[IMPL]” tag.

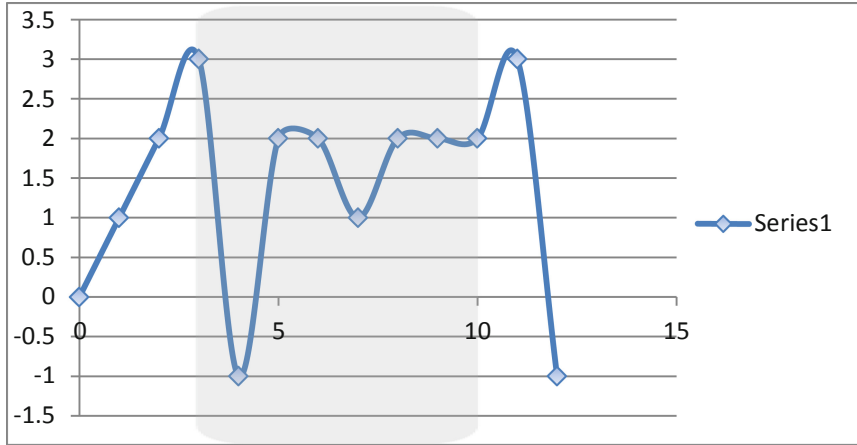


Fig. 4. “Hot spots” -Tension (shaded area between topics) in generated graphical representation producing the [IMPL] tag as output.

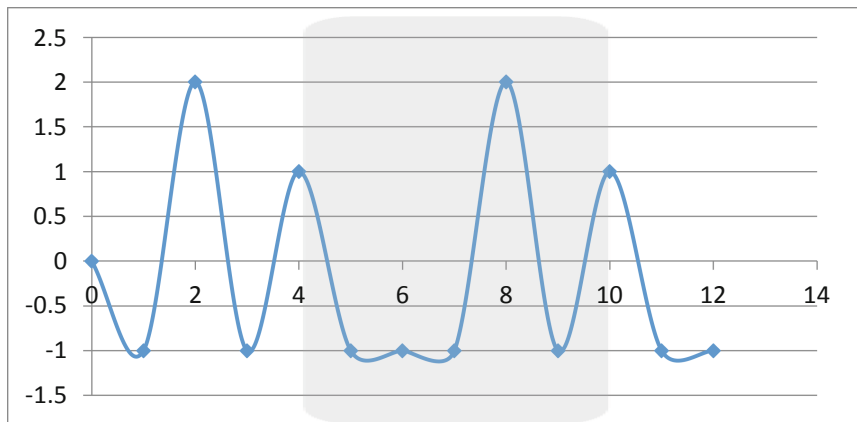


Fig. 5. “Hot spots” -Tension (shaded area between topics) in generated graphical representation with multiple “Topic Switch” relations, producing the [IMPL] tag as output.

4 Generating and Annotating Information Not Uttered in Paralinguistic Features

The generated graphic patterns allow the additional indication of any paralinguistic features influencing the content of the spoken utterances. Since paralinguistic features concern information that is not uttered, the signalization and visualization of such information plays an important role in the correct and complete transfer of the information content, in accordance to the Gricean Cooperative Principle. The Gricean Cooperative Principle is violated if the information conveyed is perceived as not complete (Violation of Quantity or Manner) or even contradicted by paralinguistic

features (Violation of Quality). Paralinguistic features may constitute pointers to information content (A. Pointer) or can be referred to as “stand-alone” information (B. Stand-Alone) [2].

The “Presence” Pointer, “Make Impression” Pointer or “Express Policy” Pointer to the Speaker’s intentions and behavior is also related to paralinguistic features. Since paralinguistic features concern information that is not uttered, the signalization of such information plays an important role in the correct and complete transfer of the information content, in accordance to the Gricean Cooperative Principle. The Gricean Cooperative Principle is violated if the information conveyed is perceived as not complete (Violation of Quantity or Manner) or even contradicted by paralinguistic features (Violation of Quality).

Paralinguistic features constituting pointers to information content (A. Pointer) may be indicated either (i) with adaptations in the transcription and/or translation (for example, the insertion of modifiers or explanatory elements) or (ii) with the insertion of a separate message or response [Message/Response] as an annotation appended to the transcription of the spoken utterance.

Paralinguistic features referred to as “stand-alone” information (B. Stand-Alone) may require the insertion of an additional utterance in the text constituting the transcription and/or translation. In this case, the insertion of a separate message or response [Message/Response] to the transcription does not correspond to a transcribed text segment but inserted as an additional feature. Therefore, for example, the raising of eyebrows with the interpretation “I am surprised” [and/but this surprises me] [2] may be indicated either as [I am surprised], as a pointer to information content (A. Pointer), or as [Message/Response: I am surprised], as a substitute of spoken information, a “stand-alone” paralinguistic feature (B. Stand-Alone).

The alternative interpretations of the paralinguistic feature (namely, “I am listening very carefully”, “What I am saying is important” or “I have no intention of doing otherwise”) [2] can be indicated with the annotations “[I am listening], [Please pay attention], [No] and [Message/Response: I am listening], [Message/Response: Please pay attention], [Message/Response: No]” respectively. The insertion of the respective type of annotation depends on whether paralinguistic feature constitute “Pointer” (A) or “Stand-Alone” (B) paralinguistic features.

Similarly, the slight raise of hand outward with the interpretation “Wait a second” [and/but wait] [2] may be either be indicated as [Stop. Wait], as a pointer to information content (A. Pointer), or as [Message/Response: Stop. Wait.], as a substitute of spoken information, a “stand-alone” paralinguistic feature (B. “Stand-Alone”). The alternative interpretations of the paralinguistic feature (namely, “Let me speak”, “I disagree with this” or “Stop what you are doing”) [2] can be indicated with the annotations “[Let me speak], [No], [Stop] and [Message/Response: Let me speak], [Message/Response: No], [Message/Response: Stop]” respectively. The insertion of the respective type of annotation depends on whether paralinguistic feature constitute “Pointer” (A) or “Stand-Alone” (B) paralinguistic features.

In the proposed framework, the interactive annotation of the previously described prosodic features is combined with the option of indicating the respective paralinguistic

features ([facial-expr: type], [gesture: type]), if applicable, and the insertion of the chosen annotations, for example “[facial-expr: eyebrow-raise]” and “[gesture: low-hand-raise]”. The insertion of the respective annotation allows the insertion/generation of the appropriate messages, according to the parameters of the language(s) and the speaker(s) concerned.

Paralinguistic features are annotated interactively with the corresponding tags and/or the chosen respective messages. In this case, the generation of the [IMPL] tag for an entire speech segment depends on the user’s evaluation of the paralinguistic features concerned. One of the intended functions of the proposed annotation is its use as an additional annotation option to existing transcription tools and speech processing applications. The annotations concern text output generated by Speech Recognition (ASR) module for pre-processing/post-processing, providing options for evaluation, (machine) translation or other processes, including Data Mining applications. The annotation can be run as an additional process or with a possible integration (as upgrade) in existing tools and systems.

In case of the interactive annotation of the paralinguistic features the [IMPL] tag is not automatically generated. This difference is related to the particularities of the information content of paralinguistic features perceived by the user (Figs. 6, 7 and 8).

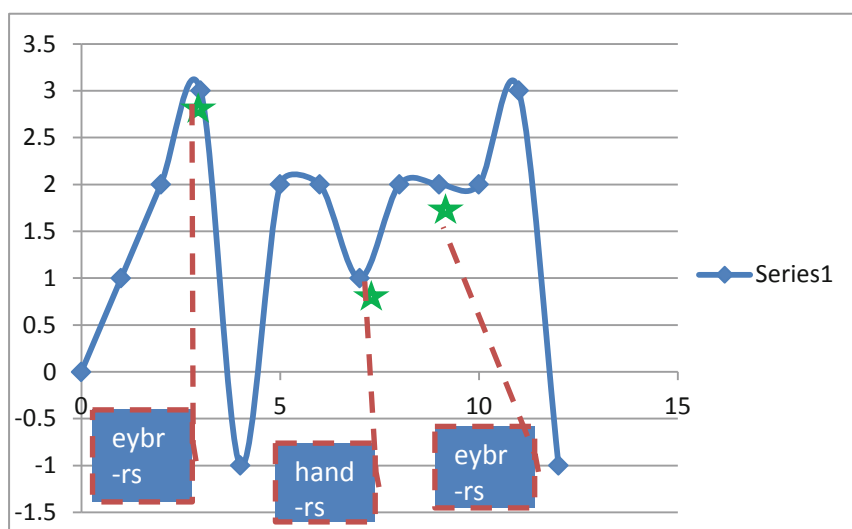


Fig. 6. Paralinguistic information (annotations) in generated graphical representation. The “[facial-expr: eyebrow-raise]” and “[gesture: low-hand-raise]” annotations depicted as “[eybr-rs]” and “[hand-rs]” respectively. The [IMPL] tag is a result of the user’s choice and evaluation.

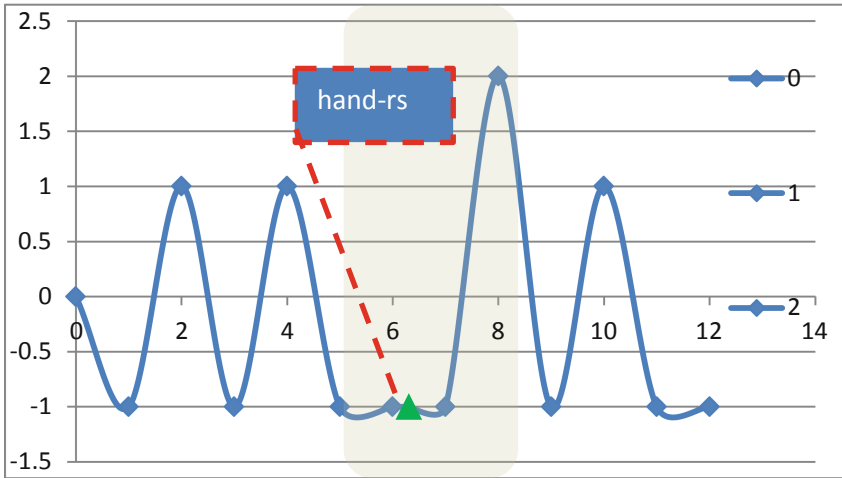


Fig. 7. “Hot spots” -Tension (shaded area between topics) and paralinguistic information (annotations) in generated graphical representation with multiple “Topic Switch” relations, producing the [IMPL] tag as output.

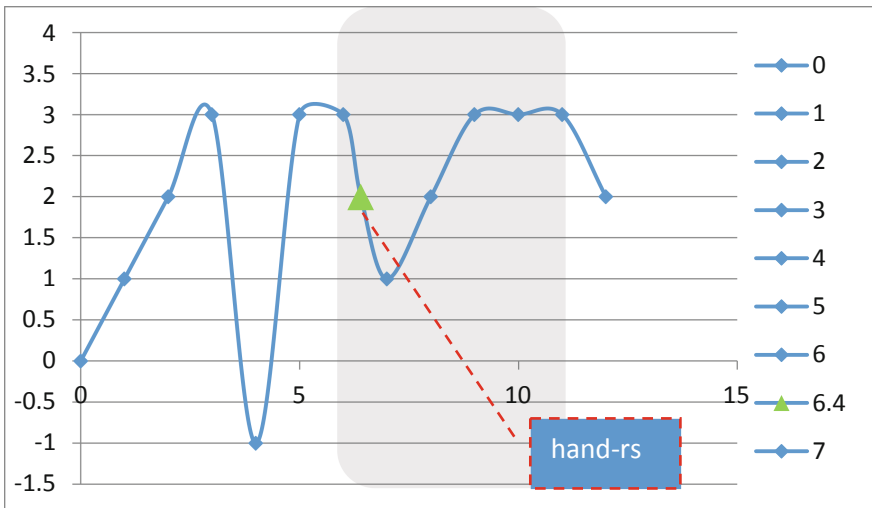


Fig. 8. “Hot spots” -Tension (shaded area between topics) and paralinguistic information (annotation) in generated graphical representation with multiple “Generalization” relations, producing the [IMPL] tag as output.

5 Conclusions and Further Research: Interface Upgrade and Empirical Data for Applications

The present application targets to assist the evaluation and decision-making process in respect to discussions and interviews in the Media (or Skype), providing a graphic representation of the discourse structure and aiming to by-pass Cognitive Bias of the user-evaluator (and/or User-Journalist). The predominate types of relations in the discourse and dialog structure, if applicable, are easily identified by the y level value around which the graphic representation is developed.

The time-frame generation of the linear structure allows the graphic representation to be presented in conjunction with the parallel depiction of speech signals and transcribed texts, a typical feature of most transcription tools. In other words, the alignment of the generated graphic representation with the respective segments of the spoken text enables a possible integration of the present application in existing transcription tools.

Furthermore, the above-described graphic representations and values enable the evaluation of the behavior of speakers-participants, allowing the identification and detection of additional, “hidden” Illocutionary Acts not restricted to “Obtaining Information Asked” or “Providing Information Asked” framework defined by the interview or discussion.

A further development and upgrading of the current interface is necessary for increasing speed and ameliorating user-friendliness. The envisioned upgrade includes the simplification of the existing menu and overall improvement of the graphical user interface (GUI).

In the present application, special focus is placed on tension in spoken political and journalistic texts as a source of empirical data both for human behaviour and for linguistic phenomena, especially when an international public is concerned and where a variety of linguistic and socio-cultural factors is included. With the visibility of all information content, including information not uttered, the proposed processing and annotation approaches may also be used for compiling empirical data for research and/or for the development of HCI- HRI Sentiment Analysis and Opinion Mining applications, as (initial) training and test sets or for Speaker (User) behavior and expectations.

References

1. Alexandris, C.: Evaluating cognitive bias in two-party and multi-party spoken interactions. In: Proceedings from the AAAI Spring Symposium, Stanford University (2019)
2. Alexandris, C.: Visualizing pragmatic features in spoken interaction: intentions, behavior and evaluation. In: Proceedings of the 1st International Conference on Linguistics Research on the Era of Artificial Intelligence – LREAI, Dalian, 25–27 October 2019. Dalian Maritime University (2010)
3. Alexandris, C.: Measuring cognitive bias in spoken interaction and conversation: generating visual representations. In: Beyond Machine Intelligence: Understanding Cognitive Bias and Humanity for Well-Being AI Papers from the AAAI Spring Symposium, Stanford University, Technical Report SS-18-03, pp. 204–206. AAAI Press, Palo Alto (2018)

4. Alexandris, C., Nottas, M., Cambourakis, G.: Interactive evaluation of pragmatic features in spoken journalistic texts. In: Kurosu, M. (ed.) HCI 2015. LNCS, vol. 9171, pp. 259–268. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-21006-3_26
5. Alexandris, C.: English, German and the international “semi-professional” translator: a morphological approach to implied connotative features. *J. Lang. Translation*. **11**(2), 7–46 (2010)
6. Arockiaraj, C.M.: Applications of neural networks in data mining. *Int. J. Eng. Sci.* **3**(1), 8–11 (2013)
7. Austin J.L.: How to do things with words. In: Urmson, J.O., Sbisà, M. (eds.), 2nd edn. University Press, Oxford Paperbacks, Oxford (1962)
8. Carlson, L., Marcu, D., Okurowski, M.E.: Building a discourse-tagged corpus in the framework of rhetorical structure theory. In: Kuppevelt, J., Smith, R.W. (eds.) *Current and New Directions in Discourse and Dialogue*. TLTB, vol. 22, pp. 85–112. Springer, Dordrecht (2003). https://doi.org/10.1007/978-94-010-0019-2_5
9. Cohen, P., et al.: Quickset: multimodal interaction for distributed applications. In: Proceedings of the 5th ACM International Multimedia Conference, Seattle, Washington, pp. 31–40. Association for Computing Machinery (ACM) (1997)
10. Cutts, M.: *Oxford Guide to Plain English*, 4th edn. Oxford University Press, Oxford (2013)
11. Floros, V., Mourouzidis, D.: *Multiple Task Management in a Dialog System for Call Centers*. Master’s Thesis, Department of Informatics and Telecommunications, National University of Athens, Greece (2016)
12. Grice, H.P.: *Studies in the Way of Words*. Harvard University Press, Cambridge (1989)
13. Grice, H.P.: Logic and conversation. In: Cole, P., Morgan, J. (eds.) *Syntax and Semantics*, vol. 3. Academic Press, New York (1975)
14. Hatim, B.: *Communication Across Cultures: Translation Theory and Contrastive Text Linguistics*. University of Exeter Press, Exeter (1997)
15. Hedderich, M.A., Klakow, D.: Training a neural network in a low-resource setting on automatically annotated noisy data. In: Proceedings of the Workshop on Deep Learning Approaches for Low-Resource NLP, Melbourne, Australia, pp. 12–18. Association for Computational Linguistics-ACL (2018)
16. Hilbert, M.: Toward a synthesis of cognitive biases: how noisy information processing can bias human decision making. *Psychol. Bull.* **138**(2), 211–237 (2012)
17. Lewis, J.R.: *Introduction to Practical Speech User Interface Design for Interactive Voice Response Applications*. IBM Software Group, San Diego (2009). USA, Tutorial T09 presented at HCI 2009
18. Liu, B.: *Sentiment Analysis and Opinion Mining*. Morgan & Claypool, San Rafael (2012)
19. Ma, J.: A comparative analysis of the ambiguity resolution of two English-Chinese MT approaches: RBMT and SMT. *Dalian Univ. Technol. J.* **31**(3), 114–119 (2010)
20. Marcu, D.: Discourse trees are good indicators of importance in text. In: Mani, I., Maybury, M. (eds.) *Advances in Automatic Text Summarization*, pp. 123–136. The MIT Press, Cambridge (1999)
21. Mourouzidis, D., Floros, V., Alexandris, C.: Generating graphic representations of spoken interactions from journalistic data. In: Kurosu, M. (ed.) HCI 2019. LNCS, vol. 11566, pp. 559–570. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-22646-6_42
22. Nass, C., Brave, S.: *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. The MIT Press, Cambridge (2005)
23. Nottas, M., Alexandris, C., Tsopanoglou, A., Bakamidis, S.: A hybrid approach to dialog input in the citizenship dialog system for consumer complaints. In: Proceedings of HCI 2007, Beijing, People’s Republic of China (2007)
24. Paltridge, B.: *Discourse Analysis: An Introduction*. Bloomsbury Publishing, London (2012)

25. Pan, Y.: Politeness in Chinese face-to-face interaction. In: *Advances in Discourse Processes Series*, vol. 67. Ablex Publishing Corporation, Stamford (2000)
26. Poria, S., Cambria, E., Hazarika, D., Mazumder, N., Zadeh, A., Morency, L.-P.: Context-dependent sentiment analysis in user-generated videos. In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Vancouver, Canada, 30 July–4 August 2017, pp. 873–883. Association for Computational Linguistics – ACL (2017). <https://doi.org/10.18653/v1/P17-1081>
27. Sacks, H., Schegloff, E.A., Jefferson, G.: A simplest systematics for the organization of turn-taking for conversation. *Language* **50**, 696–735 (1974)
28. Searle, J.R.: *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, Cambridge (1969)
29. Shah, K., Kopru, S., Ruvini, J.-D.: Neural network based extreme classification and similarity models for product matching. In: *Proceedings of NAACL-HLT 2018*, New Orleans, Louisiana, 1–6 June 2018, pp. 8–15. Association for Computational Linguistics-ACL (2018)
30. Stede, M., Taboada, M., Das, D.: *Annotation Guidelines for Rhetorical Structure*. Manuscript. University of Potsdam and Simon Fraser University, March 2017
31. Trofimova, I.: Observer bias: an interaction of temperament traits with biases in the semantic perception of lexical material. *PloSone* **9**(1), e85677 (2014)
32. Wardhaugh, R.: *An Introduction to Sociolinguistics*, 2nd edn. Blackwell, Oxford, UK (1992)
33. Williams, J.D., Asadi, K., Zweig, G.: Hybrid Code Networks: practical and efficient end-to-end dialog control with supervised and reinforcement learning. In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, Vancouver, Canada, 30 July–4 August 2017, pp. 665–677. Association for Computational Linguistics (ACL) (2017)
34. Wilson, M., Wilson, T.P.: An oscillator model of the timing of turn taking. *Psychon. Bull. Rev.* **12**(6), 957–968 (2005). <https://doi.org/10.3758/BF03206432>
35. Yu, Z., Yu, Z., Aoyama, H., Ozeki, M., Nakamura, Y.: Capture, Recognition, and Visualization of Human Semantic Interactions in Meetings. In: *Proceedings of PerCom*, Mannheim, Germany (2010)
36. Zeldes, A.: rstWeb - a browser-based annotation interface for rhetorical structure theory and discourse relations. In: *Proceedings of NAACL-HLT 2016 System Demonstrations*. San Diego, CA, pp. 1–5 (2016). <http://aclweb.org/anthology/N/N16/N16-3001.pdf>