Evaluating Cognitive Bias in Two-Party and Multi-Party Spoken Interactions

Christina Alexandris

National and Kapodistrian University of Athens
calexandris@gs.uoa.gr

Abstract

Targeting to by-pass Cognitive Bias in two-party discussions and interviews containing longer speech segments, a proposed semi-automatic procedure involves "taking the temperature" of a transcribed dialog by measuring the number of detected points of possible tension and/or conflict between speakers-participants.

Introduction

Human Computer Interaction (HCI) systems may assist in the evaluation of complex Human-Human interaction, as in the case of designed applications for journalists (Alexandris, Nottas and Cambourakis, 2015).

In spoken dialogs concerning complex interactions between speakers-participants -as in the case of spoken journalistic texts, there are aspects that can be evaluated by semi-automatic or interactive procedures, targeting to by-pass Cognitive Bias and there are aspects that can be evaluated by interactive procedures, targeting to register Cognitive Bias.

Unlike task-specific dialogs (Tung et al., 2013) and typical collaborative dialogs (Yang, Levow and Meng, 2012, Wang et al., 2013), the Speech Acts performed by one or multiple speakers-participants, often may involve complex Illocutionary Acts beyond the defined framework of the interaction. Specifically, the Illocutionary Act (Searle,1969, Austin, 1962) performed by the Speaker concerned may not be restricted to "Obtaining Information Asked" or "Providing Information Asked" in a discussion or interview: Speakers-participants may have other or additional intentions regarding their presence and their role in the discussion or interview concerned. In the spoken journalistic texts concerned, Illocutionary Acts not restricted to "Obtaining Information Asked" or "Providing Information Asked" are related to other or additional Speaker intentions. For example, a Speaker may focus in emphasizing opinion (or the policy of the network concerned) or in (purposefully) creating tension in the interview or discussion. Furthermore, a consistent avoidance of the topics addressed may indicate that the Speaker is more interested in showing a mere presence in the discussion or interview, rather than sharing any information.

The existence of additional, "hidden" Illocutionary Acts can be identified, by procedures evaluating the behavior of speakers-participants in relation to specific values and benchmarks. The presentation and calculation of these values allows the possibility of by-passing or registering Cognitive Bias. The Cognitive Bias by-passed or registered concerns primarily the evaluation of Cognitive Bias of (i) the speakers-participants concerned but may also serve for the evaluation of the Confidence Bias of (ii) the user-evaluator of the recorded and transcribed discussion or interview.

By-Passing and Registering Cognitive Bias in Multiple Speaker Discussions or in Short Speech Segments

In smaller speech segments with constant and quick change of speaker turns and with discourse structure compatible to models where each participant selects self (Wilson, 2005, Sacks, Schegloff, and Jefferson 1974), topic tracking (and topic change) allows the evaluation of speaker behavior and enables the identification of speaker's intentions and Illocutionary Speech Acts performed (Searle,1969, Austin, 1962). Topic tracking can be applied especially in short speech segments with two or multiple speakers-participants (Alexandris, 2018). The content of relatively short utterances can be summarized with the use of keywords chosen from each utterance by the user-evaluator (Alexandris, 2018), 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 (Stede, Taboada and Das, 2017, Zeldes, 2016) usually involves larger (written) texts and may not produce the required results.

The System 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 (Nottas et al., 2007, Floros and Mourouzidis, 2016). This function is similar to user-independent evaluations of spoken dialog systems (Williams, Asadi and Zweig, 2017) for by-passing User bias (Nass and Brave, 2005, Cohen, 1997). Keywords (topics) may be repeated or related to a more general concept (or global variable) (Lewis, 2009) 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) (Alexandris, 2018). 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 (Alexandris, 2018).

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 (Hilbert, 2012) 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, (Trofimova, 2014) and may differ according to political, socio-cultural and linguistic characteristics of the user-evaluator, especially if international users are concerned (Yu et al., 2010, Alexandris, 2010, Ma, 2010, Pan, 2000) due to to lack of world knowledge of the language community involved (Paltridge, 2012, Hatim, 1997, Wardhaugh, 1992). The envisioned further development of generated visual representations is their modeling in a form of graphs, similar to discourse trees (Marcu, 1999, Carlson, Marcu and Okurowski, 2001).

Evaluation and Benchmarks

The types of relations-distances between word-topics chosen by the user-evaluator are registered and counted. If the number of (a) the "Repetitions" label 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 Ds = Number of Distances and Sp = Number of Speaker turns:

- $X = Ds \leq Sp$ (calculating over 50% of "Repetitions" (Distance = 1, value "1")) or "Topic Switches" (Distance = 4, value "0").
- $X = Ds > Sp \times Gen$ ($Gen = Sp \times 3 \div 2$) (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)".

By-Passing and Registering Cognitive Bias in Two-Party Discussions and Interviews or in Long Speech Segments

The further development of the database containing registered spoken interaction for determining and evaluating Cognitive Bias in spoken journalistic texts (Alexandris, 2018) involved the processing of discussions and interviews containing larger speech segments. Similarly to the above-described multiple speaker discussions and in short speech segments, 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.

In two-party discussions and interviews containing longer speech segments, the discourse structure is more compatible to turn-taking in "push-to-talk conversations", with a strict protocol in managing the interview or discussion and turn-taking (Taboada, 2006). In this case, speakers-participants usually not have the liberty of modifying or changing the topic, resulting to the strategy of topic tracking being insufficient for the identification of speaker's intentions. In larger speech segments mostly occurring in interviews with a strict protocol and a set of predefined topics, automatic Rhetorical Structure Theory (RST) analysis procedures (Stede, Taboada and Das, 2017, Zeldes, 2016) can be performed in the transcribed text, with the condition that the speaker is allowed sufficient time to elaborate on the topic in question. The extent to which automatic RST analysis procedures can be executed in the transcribed text indicates the degree of collaborative interaction between the speakers-participants, especially from the journalist-interviewer (referred to as Speaker 1), since the speaker-participant is allocated enough time to elaborate and/or argument on the topic concerned.

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 signalizing and counting the points of possible tension and/or conflict between speakers-participants henceforth referred to as "hotspots". The signalization of "hotspots" is based on the violation of the Quantity, Quality and Manner Maxims of the Gricean Cooperativity Principle (Grice, 1975). Cognitive Bias is registered by comparing content of the Speaker turns in the signalized "hotspots" and assigning a respective value.

By-Passing Cognitive Bias: Automatic Signalization of "Hotspots" and the Gricean Cooperativity Principle

Targeting to by-pass Cognitive Bias in two-party discussions and interviews containing longer speech segments, a proposed semi-automatic procedure involves "taking the temperature" of a transcribed dialog by measuring the number of detected points of possible tension and/or conflict between speakers-participants. These points are henceforth, referred to as "hot spots" and concern in speech segments where there is a recognition of speaker turns, namely a switch between Speaker 1 and Speaker 2 by the Speech recognition module of the transcription tool. 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, (Cutts 2013) 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 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 (Grice, 1975). These conditions are the following:

- (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). These forms of adjectival or adverbial phrases are detectable with a POS Tagger (for example, the Stanford POS Tagger.
- (2) Reference to the interaction itself and to its participants with negation. In one or in both speakers' utterances, the subject of the sentence containing the negation is "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"). This applies 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"). The different forms of negation are detectable with a POS Tagger. The respective words and word categories may constitute a small set of entries in a specially created lexicon or may be retrieved from existing databases or WordNets .
- (3) Prosodic emphasis and/or Exclamations. (a) Exclamations include expressions such as such as "Look", "Wait" and "Stop". As in the above-described case (2), the respective words and word categories may constitute a small set of entries in a specially created lexicon or may be retrieved from existing databases or WordNets. (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.

In the case of 1a, 1b and 1c, there is extra information added to the basic content of the utterance consisting the necessary information required to fulfil the Gricean Cooperative Principle in respect to the Maxim of Quantity. ("Do not make your contribution more informative than is required "). Here, the Speaker violates the Maxim of Quantity in the Gricean Cooperative Principle. In the case of 2a and 2b, the Speaker perceives a violation of the Gricean Cooperative

Principle by the previous Speaker. In particular, the content of the speaker's utterance is not limited to the current topic in question but refers to the dialog itself, mostly functioning as a comment. Specifically, 2a and 2b imply a 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") (Grice, 1975) and/or in respect to the Maxim of Manner (Submaxim 2. "Avoid ambiguity") (Grice, 1975) in the utterance of the previous Speaker. In other words, in 2a and 2b, the Speaker considers the content of the previous Speaker's utterance to be unacceptable, ambiguous, false or controversial.

The number of automatically signalized "hot spots" indicates the degree in which discussions and interviews containing larger speech segments constitute dialog with many points of tension and/or conflict. The average time of discussions and interviews containing larger speech segments in the Media is 30 to 45 minutes (30-45 mins). A typical example of a dialog with many detected points of possible tension and/or conflict between speakers-participants is an approximately 32 minute long interview with seven (7) registered "hot spots" (BBC (British Broadcasting Corporation): HARDtalk interview by journalist Stephen Sackur on 16th of April 2018). One or both speakers' utterances may display two or more of features (1), (2) and (3).

Evaluation and Benchmarks

The benchmark for evaluating a remarkable degree of tension in a discussion is signalized by multiple "hotspots" detected and not sporadic occurrences of "hotspots". Thus, the number of 1-2 "hotspot" occurrences in longer speech segments in question (30-45 mins) signalizes a low degree of tension. A remarkable degree of tension in a 30-45 minute discussion or interview is related to a number of at least 4 detected "hotspots" (where the number of 3 hotspots constitutes a marginal value). Considering the above, the benchmark for evaluating a remarkable degree of tension concerns the calculation of the time of discussion / interview in the Media (for example, 35 mins) and the number of signalized "hot spots" (SPEECH SEGMENT-count) in Speaker turns. The defined benchmark for evaluating Speaker behavior is the number of minutes divided by the number of identified speech segments signalized as "hot spots" which should contain a single digit number, if the above-described minimal number of at least 4 detected "hotspots" is calculated. For example, the acceptable values are "8.75", "7" or, ideally, "5" (for a file of 35 minutes) versus "17.5" or "11.6" (for a file of 35 minutes). Interactions with Illocutionary Acts beyond the predefined framework of the dialog-discussion (with two speakers –participants and long speech segments) are based on the detected points of possible tension and/or conflict indicated by the following benchmark, where

Y = wav file length in minutes divided by (÷) the number of "hot spot" signalized speech segments:

- $Y < 10$.
- Example: File length = 35 mins, 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)".

Registering Cognitive Bias: Interactive Comparison of Speaker Turns

The registration of Cognitive Bias concerns the comparison of the actual content of the pair of utterances of Speaker 1 and Speaker 2 in the signalized "hot spots". As stated above, the automatically signalized "hot spots" are extracted to a separate template for interactive processing, where the "hot spot" utterances of both speakers are compared. If the last 60 words (approximately 1 -3 sentences, with average sentence length of 15-20 words, (Cutts 2013) of the first speaker's utterance contain at least two of the above-described features (1), (2) and (3), the Quantity and Manner Maxims of the Gricean Cooperative Principle (Grice, 1975) are violated. Specifically, Cognitive Bias is registered by comparing content of the Speaker turns in the signalized "hotspots" and assigning the following respective values:

- (a) Each "hot spot" is marked with a (1,1) if both speakers' utterances are considered equally non-collaborative.
- (b) If this is the case for one of the two speakers, in particular, Speaker 1, the "hot spot" is marked with a (1,0) for Speaker 1 (in this case, the journalist-reporter). In this case, the style of question or statement uttered is not considered acceptable- contains features violating the Gricean Cooperative Principle - in respect to the Maxim of Manner or the Maxim of Quantity (Irony) or in respect to the Maxim of Quality (content is considered false ("F").
- (c) If this is the case for Speaker 2, the "hot spot" is marked with a (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.
- (d) If a "hot spot" speech segment is evaluated by the User not as a point of possible tension and/or conflict between speakers-participants, the false "hot spot" is marked with a (0,0) for both Speakers.

Evaluation and Benchmarks

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: Evaluation of Speaker Behavior (Speaker 1 is less collaborative than Speaker 2).
- SPEAKER1: (1), (1), (1), (0), (1).
- SPEAKER2: (0), (0), (1), (1), (0).
- File length: 35 mins: 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)".

Conclusions and Further Research

By-passing and registering Cognitive Bias in HCI systems assisting in the evaluation of Human-Human interaction involves both automatic and interactive procedures. Interactive topic tracking in the dialog structure and automatic "hot spot" generation involving points of tension and/or conflict contribute to an evaluation of speakers-participants behavior and intentions during the interaction.

The behavior and Cognitive Bias of (i) speakers-participants is evaluated in relation to the values of the "Relevance (X)", "Tension (Y)" and "Collaboration (Z)" benchmarks. However, the same benchmarks may be used for evaluating the Cognitive Bias- Confidence Bias of (ii) the user-evaluator of the recorded and transcribed discussion or interview.

Spoken dialogs concerning complex interactions between speakers-participants are not limited to spoken journalistic texts. As the variety and complexity of spoken HCI applications increases, Speech Acts performed by one or multiple users-participants, even by the System itself, often may involve Illocutionary Acts beyond the predefined framework of a task-oriented dialog, especially in systems with emotion recognition, virtual negotiation, psychological support or decision-making.

References

Alexandris, C. 2018. Measuring Cognitive Bias in Spoken Interaction and Conversation: Generating Visual Representations. In: Beyond Machine Intelligence: Understanding Cognitive Bias and Humanity for Well-Being AI. In *Proceedings from the AAAI Spring Symposium*, Stanford University, 204-206 Technical Report, SS-18-03, Palo Alto, CA: AAAI Press.

Alexandris, C. ; Nottas, M. ; and Cambourakis, G. 2015. Interactive Evaluation of Pragmatic Features in Spoken Journalistic Texts. In Kurosu, M. ed., Human-Computer Interaction, Users and Contexts, *LNCS Lecture Notes in Computer Science*, Vol. 9171: 259-268, Heidelberg, Germany: Springer.

Alexandris, C. 2010. English, German and the International "Semi-professional" Translator: A Morphological Approach to Implied Connotative Features. *Journal of Language and Translation*, September 2010, Vol. 11, 2, Sejong University, Korea: 7- 46.

Austin J. L. 1962. *How to Do Things with Words*. Urmson. J.O. and Sbisà, M. eds., 2nd edition., 1976, Oxford, UK: Oxford University Press.

Carlson, L.; Marcu, D. ; and Okurowski, M. E. 2001. Building a Discourse-Tagged Corpus in the Framework of Rhetorical Structure Theory. In *Proceedings of the 2nd SIGDIAL Workshop on Discourse and Dialogue*, Eurospeech 2001, Denmark, September 2001, ACL Anthology, W01-16.

Cohen, P. ; Johnston, M. ; McGee, D. ; Oviatt, S. ; Pittman, J. ; Smith, I. ; Chen, L. ; and Clow, J. 1997. Quickset: Multimodal Interaction for Distributed Applications. In *Proceedings of the 5th ACM International Multimedia Conference*, 31-40, New York, NY: ACM Digital Library.

Cutts, M. 2013. *Oxford Guide to Plain English*. 4th edition., Oxford, UK: Oxford University Press.

Du, J. ; Alexandris, C. ; Mourouzidis, D. ; Floros, V. ; and Iliakis, A. 2017. Controlling Interaction in Multilingual Conversation Revisited: A Perspective for Services and Interviews in Mandarin Chinese. In Kurosu, M. ed., *Lecture Notes in Computer Science* LNCS 10271: 573–583, Heidelberg, Germany: Springer.

Floros, V. and Mourouzidis, D. 2016. Multiple Task Management in a Dialog System for Call Centers. Master's Thesis, Department of Informatics and Telecommunications, National University of Athens, Greece.

Grice, H.P. 1975. Logic and conversation. In: Cole, P., Morgan, J. eds., *Syntax and Semantics*, Vol. 3. Academic Press, New York.

Hatim, B. 1997. *Communication Across Cultures: Translation Theory and Contrastive Text Linguistics*. Exeter, UK: University of Exeter Press.

Hilbert, M. 2012. Toward a Synthesis of Cognitive Biases: How Noisy Information Processing Can Bias Human Decision Making. *Psychological Bulletin*, Vol 138(2), Mar 2012: 211-237.

Lewis, J.R. 2009. Introduction to Practical Speech User Interface Design for Interactive Voice Response Applications, IBM Software Group, USA, Tutorial T09 presented at HCI 2009 San Diego, CA, USA

Ma, J. 2010. A comparative analysis of the ambiguity resolution of two English-Chinese MT approaches: RBMT and SMT. *Dalian University of Technology Journal*, 31(3): 114-119.

Marcu, D. 1999. Discourse trees are good indicators of importance in text. In Mani, I. and Maybury, M. (eds), *Advances in Automatic Text Summarization*, Cambridge MA, The MIT Press: 123-136.

Nass, C. and Brave, S. 2005. *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. Cambridge MA: The MIT Press.

Nottas, M. ; Alexandris, C. ; Tsopanoglou, A. ; and Bakamidis, S. 2007. A Hybrid Approach to Dialog Input in the CitzenShield Dialog System for Consumer Complaints. In *Proceedings of HCII 2007*, Beijing, Peoples Republic of China.

Paltridge, B. 2012. *Discourse Analysis: An Introduction*. London, UK: Bloomsbury Publishing.

Pan, Y. 2000. Politeness in Chinese Face-to-Face Interaction. *Advances in Discourse Processes Series* Vol. 67, Stamford, CT: Ablex Publishing Corporation.

Sacks, H. ; Schegloff, E. A. ; and Jefferson, G. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, Vol. 50: 696-735.

Searle J. R. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge, MA: Cambridge University Press.

Taboada, M. 2006. Spontaneous and non-spontaneous turn-taking *Pragmatics*, Vol.16 (2-3): 329-360.

Stede, M. ; Taboada, M., ; and Das, D. 2017. Annotation Guidelines for Rhetorical Structure. Manuscript. University of Potsdam and Simon Fraser University. March 2017.

Trofimova I. 2014. Observer Bias: An Interaction of Temperament Traits with Biases in the Semantic Perception of Lexical Material. *PLoS ONE* 9(1): e85677.

Tung, T. ; Gomez, R. ; Kawahara, T. ; and Matsuyama, T. 2013. Multi-party Human-Machine Interaction Using a Smart Multimodal Digital Signage. In Kurosu, M. ed., *Human-Computer Interaction. Interaction Modalities and Techniques, Lecture Notes in Computer Science*, Vol. 8007, 2013, 408-415, Heidelberg, Germany: Springer.

Wang, H. ; Gailliot, A. ; Hyden, D. ; and Lietzenmayer, R. 2013. A Knowledge Elicitation Study for Collaborative Dialogue Strategies Used to Handle Uncertainties in Speech Communication While Using GIS. In Kurosu, M. ed., *Human-Computer Interaction, Lecture Notes in Computer Science* 8007, 135-146, Heidelberg, Germany: Springer.

Wardhaugh, R. 1992. *An Introduction to Sociolinguistics*. 2nd edition. Oxford, UK: Blackwell.

Williams, J.D. ; Asadi, K. ; and Zweig, G. 2017. 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, July 30 - August 4, 2017, 665–677, Association for Computational Linguistics -ACL.

Wilson, K. E. 2005. An oscillator model of the timing of turn-taking. *Psychonomic Bulletin and Review* 2005:12 (6): 957-968.

Yang, Z. ; Levow G.A. ; and Meng F. H. 2012. Predicting User Satisfaction in Spoken Dialog System Evaluation With Collaborative Filtering. *IEEE Journal of Selected Topics in Signal Processing*, Vol. 6, Issue: 8, Dec. 2012: 971 – 981.

Yu, Z. ; Yu, Z. ; Aoyama, H. ; Ozeki, M.; and Nakamura, Y. 2010. Capture, Recognition, and Visualization of Human Semantic Interactions in Meetings. In *Proceedings of PerCom, Mannheim, Germany*, 2010.

Zeldes, A. 2016. "rstWeb - A Browser-based Annotation Interface for Rhetorical Structure Theory and Discourse Relations". In *Proceedings of NAACL-HLT 2016 System Demonstrations*, San Diego, CA, 1-5, ACL Anthology, NAACL-HLT 2016.