On the Role of Communicative Structure in Read Aloud Applications for the Elderly

Mónica Domínguez
University Pompeu Fabra
Barcelona, Spain
monica.dominguez@upf.edu

Alicia Burga
University Pompeu Fabra
Barcelona, Spain
alicia.burga@upf.edu

Mireia Farrús
University Pompeu Fabra
Barcelona, Spain
mireia.farrus@upf.edu

Leo Wanner
Catalan Institute for Research and Advanced Studies and
University Pompeu Fabra
Barcelona, Spain
leo.wanner@upf.edu

ABSTRACT

Conversational technologies that assist elderly people need to adapt to common disabilities in old age. Visual, hearing and even more so cognitive impairments pose serious difficulties for our seniors to handle a standard conversation with a human. Understanding a virtual agent may be ever harder. In this case, communicative strategies are key to adapt the virtual agent to the needs of elderly users. This paper addresses the role of the communicative structure for expressive speech prosody, which is known to be crucial for better speech comprehension. It reports on efforts to improve prosody within a text-to-speech system based on one aspect of the communicative structure, namely thematicity. The work has been implemented as an application in a social virtual agent, KRISTINA, which reads aloud news articles upon request for elderly users in German.

CCS CONCEPTS

• **Social and professional topics → Seniors**;

KEYWORDS

intelligent conversational agents, geriatric applications, communicative structure, thematicity, prosody, text-to-speech, human-machine interaction

1 INTRODUCTION

In the last decades, conversational interfaces involving text-to-speech (TTS) applications have improved expressiveness and overall naturalness to a reasonable extent. Conversational features, such as speech acts, affective states and Information Structure have been instrumental to derive more expressive prosodic contours. However, synthetic speech is still perceived as monotonous, when a text that lacks those conversational features is read aloud in the interface, i.e., when it is fed directly to the TTS application. If users of the conversational interface furthermore have some impairments, as it is usually the case with elderly people using assisting technologies, it is paramount to adapt the conversational agent's speech to guarantee the communication flow of the interaction, and thus improve the acceptance of the agent by the user. This adaptation requires advanced functionalities that usually involve several areas of expertise. In this paper, we present how theoretical linguistics can be used in computational approaches to achieve a more fine-grained communicative interaction adapted to the elderly.

Virtual agents with human interaction capabilities have a large potential for the exploration of such user-oriented advanced functionalities. We work with KRISTINA. KRISTINA is a Knowledge-Based Information Agent with Social Competence and Human Interaction Capabilities [32]. KRISTINA interacts with the user in different scenarios. One of these scenarios consists in reading the newspaper to elderly people with eyesight impairments. This target audience requires a varied range of expressiveness in the synthetic voice, which state-of-the-art text-to-speech (TTS) applications usually lack, especially when processing long monologue discourse.

This paper discusses the role of the Information (or Communicative) Structure–prosody interface for reading aloud applications, and scratches the surface of the theoretical framework behind this interface. The discussion is based upon the authors' implementation of a thematicity-based prosody module that enriches raw texts extracted from news with communicative information with the goal to achieve a more expressive reading for targeted elderly users.[1] The aim is to analyze syntactic and Information Structure, and then use high-level linguistic features derived from the analysis to generate more expressive prosody in the synthesized speech. The proposed methodology encompasses a modular pipeline consisting of (1) a tokenizer, (2) a syntactic parser, (3) a theme/rheme parser, and (3) an SSML prosody tag converter. The implementation has been tested in an experimental setting for German, using web-retrieved news articles.

The rest of the paper is structured as follows. Section 2 introduces the motivation and background of this work. In Section 3, we dive into the theoretical grounds that support the proposed computational model from a linguistic perspective. Then, Section 4 sketches how this model has been implemented within the context of KRISTINA. Finally, conclusions are drawn in Section 5.

2 MOTIVATION AND BACKGROUND

The way information is formally packaged in a sentence, known as "Information Structure", has been a fruitful field of research in linguistic studies to better understand how communication is

[1] Such an application may also be handy for other users, not only elderly.

produced and perceived. Information Structure is a wide term and its study usually involves various linguistic dimensions in connection with how content is packaged, hence its interfaces at least semantics, syntax and prosody.

Different linguistic schools have long stated that Information Structure, and, in particular, the dichotomy referred to as *theme–rheme* [17], *given–new* [28], or *topic–focus* [16] is related to intonation.[2] Moreover, prosody structure on the grounds of thematicity partitions plays a key role in the understanding of a message [8]. Empirical studies in different languages provide evidence that when thematicity and prosody are appropriately put together, comprehension of the message is positively affected (cf., e.g., [24] for German and [31] for Catalan). Several works also show that a correlation between thematicity and beat gestures, which are an important non-verbal "prosodic" means to mark rythm and to "accentuate speech" [5], improves discourse recall and comprehension [18, 20]. Therefore, there is reason to assume that a conversational application considering the notions of content packaging by means of the relation between thematicity and prosody will benefit from the same advantages as in natural conversation environments. Most of all, conversational avatars in applications for children in educational settings [26], applications for those with special needs [23] as well as for the elderly [25, 33] and, in particular, for those with cognitive impairments [34], would greatly benefit from such a communicatively-oriented improvement.

On the other hand, expressive speech that uses a varied range of prosodic cues (variation in fundamental frequency, speech rate and intensity) is often regarded as more understandable and communicative. However, previous attempts to implement the concepts of Information Structure in text-to-speech (TTS) applications are rather scarce [19, 27]. Moreover, it is usually a simple binary theme–rheme structure what is being tested in short sentences. A more fine-grained analysis of thematicity structure, as defined by Mel'čuk [22] has been proved to yield better results to predict a wider variety of prosodic contours, which are furthermore perceived as more natural when implemented in a TTS application; see, e.g. [11, 15].

3 COMMUNICATIVE STRUCTURE

Despite the great efforts along the years for defining communicative notions, studies on Information Structure have remained within the field of theoretical linguistics. These studies sometimes explore different linguistic phenomena in relation to Information Structure (e.g., discourse, dialog, anaphora, and co-reference). The Communicative Structure within the Meaning-Text Theory (MTT) comes to cope with some of the limitations other theories on Information Structure have, as this representation is devised in the context of a theoretical production-oriented linguistic model, which is described in what follows.

3.1 A Theoretical Framework for Computational Linguistics

The Meaning-Text Theory proposes a framework for language analysis and generation suitable for Natural Language Processing (NLP) applications [4]. In particular, the Meaning-Text Theory Model [21] distinguishes different levels of representation. These levels are sequentially mapped from an unordered semantic representation (SemR) through a dependency tree structure of the Syntactic Representation (SyntR) and linearized chain of lexemes onto the Morphological Representation (MorphR) to get to the ordered string of phonemes at the Phonetic Representation (PhonR). Starting from SyntR and until PhonR, there is a subdivision into deep and surface representations.

The SemR includes four structures: (1) the Semantic Structure (SemS), which is a predicate-argument (meaning) structure of the message; (2) the Semantic Communicative Structure (SemCommS), which consists of a representation of the communicative intention of the speaker; (3) the Rhetorical Structure (RhetS), which encodes the artistic intentions and stylistic decisions of the speaker (irony, humorous, etc.); and (4) the Referential Structure (RefS), which specifies real-world referents for semantic configurations. The SemCommS superimposes on the SemS the communicative properties of the meaning of the sentence to be synthesized rather than the communicative properties of the sentence itself.[3] Consequently, the functions of SemCommS are:

- organizing initial meaning into a message;
- ensuring coherence of the text of which the sentence under synthesis is supposed to be a part;
- reducing periphrastic potential of the initial SemS, specifying more precisely the meaning.

In other words, the same abstract Semantic Structure can be shared by a given set of sentences, and it is by means of the SemCommS that these sentences are distinguished at subsequent levels (namely, SyntR, MorphR and PhonR). Figure 1 sketches the common SemS of sentences from (1a) to (1d) taken from [22].

(1a) *John met the doctor at the airport.*
(1b) *The doctor was met at the airport by John.*
(1c) *The airport was where John met the doctor.*
(1d) *It was John who met the doctor at the airport.*

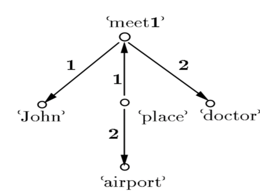

Figure 1: Shared SemS of examples (1a–1d) from [22].

The Deep Syntactic Structure (DSyntS), which may already reflect some of the SemCommS features, is the central component of the Deep-Syntactic Representation (DSyntR).[4] Consider, for illustration, the DSyntS's of sentences (1a) (Figure 2) and (1d) (Figure 3). They show how SemCommS determines the different resulting

[2]In our work, we use the first denotation, i.e., *theme–rheme* or *thematicity*. 'Theme' marks what a sentence is about, and 'rheme' what is said about the theme.

[3]In general linguistics, the term 'communicative' is usually linked to the idea of 'communicative competence' and refers to concepts related to the study of pragmatics; see the definition of 'linguistic competence' and 'performance' by Chomsky [7].

[4]Apart from DSyntS, DSyntR includes, in its turn, three further components: Deep-Syntactic Communicative Structure, Deep-Syntactic Anaphoric Structure and Deep-Syntactic Prosodic Structure (which represents semantically conditioned prosodies).

dependency trees. The communicative subject (Theme) may coincide or not with the semantic subject (Actor) and syntactic subject (Synt-Subject), as represented in Table 1. This underlines the idea that CommS is a distinct dimension.

Table 1: Communicative, semantic and syntactic subjects in examples (1a) and (1d) from [22].

(1a)	*John*	*met*	*the doctor*	*at the airport*
SemS	Actor			
SyntS	Synt-Subject			
CommS	Theme			
(1d)	*The doctor*	*was met*	*at the airport*	*by John*
SemS				Actor
SyntS	Synt-Subject			
CommS	Theme			

In a nutshell, CommS is part of the SemR and DSyntR of individual sentences. The communicative organization of text is not covered by CommS, it rather accounts for the structure of the so-called propositional content. Going back to example (1) taken from [22], the set of sentences may seem fully synonymous, but only (1a) is an appropriate reply to D1, whereas (1d) better suits D2:

D1 - *Nobody saw the doctor last night?*
- *John met him at the airport.*

D2 - *Ask John.*
- *Why John?*
- *It was John who met the doctor at the airport.*

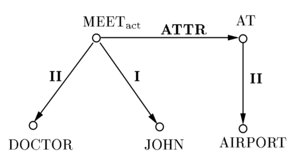

Figure 2: DSyntS from example (1a) from [22].

CommS is composed of eight distinct dimensions: 'thematicity', 'givenness', 'focalization', 'perspective', 'emphasis', 'presupposedness', 'unitariness' and 'locutionality'. As CommS characterizes the meaning of the sentence and the sentence itself, it is, consequently, modeled at the semantic level, to be propagated then to the deep-syntactic and surface-syntactic levels of the linguistic description. Note that givenness, which is often treated as synonymous to thematicity, is in Mel'čuk's communicative structure theory a distinct dimension from thematicity. According to Mel'čuk [22], the thematicity of the initial SemS has to do with psychologically motivated choices of the speaker, who decides that he/she wants to communicate some specific information (i.e., the rheme) concerning some specific item (i.e., the theme), and thereby makes the addressee follow him. In Mel'čuk's words: "The Sem-Thematicity is thus a SPEAKER-ORIENTED Comm-category."

In the following section, we sketch Mel'čuk's definition of thematicity, which is the dimension considered in previous work when the correspondence of the Information Structure with prosody is discussed.

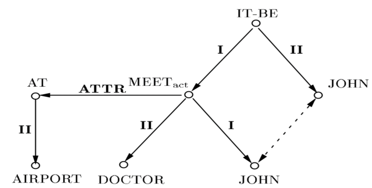

Figure 3: DSyntS from example (1d) taken from [22].

3.2 Thematicity

In contrast to Information Structure models that propose a partition of sentences into a theme and a rheme, Mel'čuk [22] argues in the context of the Meaning–Text Theory for a tripartite hierarchical division ('theme', 'rheme', and 'specifier' –the element which sets the utterance's context) within propositions that further permits embeddedness of communicative spans; consider (1) for illustration of hierarchical thematicity (annotated following the guidelines established in [2]) of the sentence *Ever since, the remaining members have been desperate for the United States to rejoin this dreadful group.* A total of five partitions are identified, including three spans at level 1, a specifier (SP1), theme (T1) and rheme (R1), and two embedded spans at level 2 in the rheme, a theme (T1(R1)) and a rheme (R1(R1)).[5]

(1) [*Ever since,*]$_{\text{SP1}}$ [*the remaining members*]$_{\text{T1}}$ [*have been desperate* [*for the United States*]$_{\text{T1(R1)}}$ [*to rejoin this dreadful group.*]$_{\text{R1(R1)}}$]$_{\text{R1}}$

A hierarchical thematicity structure of this kind has been shown to correlate better with ToBI [1, 29] labels than binary flat thematicity [10, 11]. Such a correlation still does not solve the problem of a one–to-one mapping between a specific intonation label (e.g., H*) to a static acoustic parameter (e.g., an increase of 50% in fundamental frequency). This is one of the reasons why we propose an implementation using a more varied range of automatically derived prosodic cues based on hierarchical thematicity spans, as described in what follows.

4 AUTOMATIC GENERATION OF THEMATICITY-BASED PROSODY IN KRISTINA

In the use case of KRISTINA as social companion for the elderly, the scenario of reading the newspaper involves a dialogue interaction between the user (U) and KRISTINA (K). U requests K to read the newspaper and K prompts U to pick up a piece of news. Upon reading of the title, the system retrieves the selected text, which is sent to the pipeline sketched in Figure 4. The pipeline tests the formal representation of the Communicative Structure, in particular of thematicity, proposed by Mel'čuk [22]. In the context of the conversational agent KRISTINA, text coming from a web-retrieved service is processed in the pipeline before it arrives to the TTS engine.

The proposed pipeline in Figure 4 includes four modules:

[5] As more than one thematicity span may exist within the same proposition, abbreviations include a number (e.g., 'SP1') that indicates the number of occurrences at each level (e.g., 'SP2' would be the second specifier in a specific thematicity level).

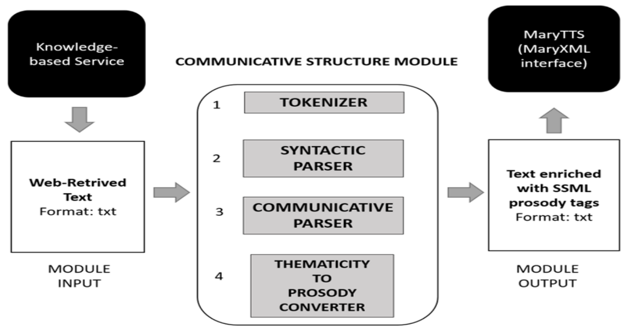

Figure 4: Communicative generation pipeline.

(1) **Tokenizer**: Splits the text into sentences and words. Punctuation marks are also tokenized as the syntactic parser requires that.
(2) **Syntactic parser**: An off-the-shelf parser [3], which is trained on the TIGER Penn Treebank [6] and which outputs a fourteen-columned CoNLL file.[6]
(3) **Communicative parser**: Derives using rules hierarchical thematicity labels from syntactic structure. It outputs a CoNLL file with an added column for communicative structure (i.e., the output CoNLL has fifteen columns).
(4) **SSML prosody converter**: Converts the thematicity spans derived by the communicative parser to SSML spans and assigns a variety of prosody tags to each span. This module is based on the tool presented in [13].

The correspondence between hierarchical thematicity and prosody is presented in terms of variations of referent SSML[7] [30] prosody tag values involving fundamental frequency (F0), speech rate (SR) and insertion of breaks.

5 CONCLUSIONS

Theoretical studies on the Information Structure-prosody interface have stated for some time that there is a correspondence between how the linguistic content is structured communicatively and how intonation is used in human speech to convey that content. In previous work, this correspondence (in particular, the relationship between hierarchical thematicity and prosodic variation) has been brought to the foreground from an empirical perspective in the context of expressive speech generation. Corpus-based experiments and data-driven implementations [12–15] supported initial expectations on the potential of the Information Structure–prosody interface applied to speech technologies. The use of this potential is an initial step ahead in communicative approaches for prosody generation within TTS/CTS applications that is one of the key aspects for a next generation of more expressive conversational virtual agents.

The implementation described above contributes in several aspects to the state of the art: (i) a formal description of hierarchical thematicity is used; (ii) a communicative parser that automatically derives thematicity labels is introduced; and (iii) a platform for prosody testing in TTS applications is demonstrated. Evaluation shows that the thematicity-based prosody enrichment is perceived as more expressive than the default TTS output. Expressiveness was assessed by means of a perception test using a Mean Opinion Score (MOS) with a 5-point Likert scale (LS): 1-bad, 2-poor, 3-fair, 4-good, and 5-excellent. Average results for the tested sentences proved that the automatic prosody modifications (LS = 3.30) achieve statistical significance at $p < 0.05$ compared to the default score (LS = 3.01). All in all, this study pivots the transition from theoretical work on the IS–prosody interface to the integration of thematicity-based prosody enrichment to achieve more expressive synthesized speech. Future work is aimed at exploring other dimensions of communicative structure like emphasis and foregroundedness within the framework that has been discussed.

Research carried out so far in this direction [9, 15] is a proof of concept of the applicability of the Information Structure–prosody interface in speech synthesis, but there are many issues that remain unexplored. For now, only thematicity at the sentence level has been tested. Other dimensions of the communicative structure (like givenness and focus, as defined by Mel'čuk [22]) may also have a strong correspondence with prosody. Corpora need to be compiled in order to continue looking into this field from an empirical perspective; see e.g. [14]. With respect to prosody, an implementation with SSML tags does not suffice to address the requirements for prosody modeling in a pre-processing stage for TTS applications. Therefore, closer insights into how to model prosody to reflect better the communicative structure of a text also need to be investigated.

Given the relevant role of the Information Structure–prosody interface in human communication, it seems reasonable that next generation conversational agents face new challenges in adopting communicatively-oriented models. In this paper, we have introduced some basic concepts on the theoretical framework behind an implementation of a hierarchical thematicity model as well as an overview of the research carried out so far in this area in its correspondence to prosody.

[6]Details about the CoNLL format are provided in http://universaldependencies.org/docs/format.html
[7]SSML stands for Speech Synthesis Markup Language: details about this convention can be found in https://www.w3.org/TR/speech-synthesis11/

REFERENCES

[1] M. E. Beckman, J. B. Hirschberg, and S. Shattuck-Hufnagel. 2004. The Original ToBI System and the Evolution of the ToBI Framework. In *Prosodic Models and Transcription: Towards Prosodic Typology*, S.A. Jun (Ed.). Oxford University Press, 9–54.

[2] B. Bohnet, A. Burga, and L. Wanner. 2013. Towards the Annotation of Penn TreeBank with Information Structure. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*. Nagoya, Japan, 1250–1256.

[3] B. Bohnet and J. Nivre. 2012. A Transition-Based System for Joint Part-of-Speech Tagging and Labeled Non-Projective Dependency Parsing. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL '12)*. Jeju Island, Korea, 1455–1465.

[4] B. Bohnet and L. Wanner. 2010. Open Source Graph Transducer Interpreter and Grammar Development Environment. In *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC)*. European Language Resources Association (ELRA), Valletta, Malta.

[5] E. Bozkurt, Y. Yemez, and E. Erzin. 2016. Multimodal analysis of speech and arm motion for prosody-driven synthesis of beat gestures. *Speech Communication* 85 (12 2016), 29–42. https://doi.org/10.1016/J.SPECOM.2016.10.004

[6] S. Brants, S. Dipper, P. Eisenberg, S. Hansen, E König, W. Lezius, C. Rohrer, G. Smith, and H. Uszkoreit. 2004. TIGER: Linguistic Interpretation of a German Corpus. *Journal of Language and Computation* 2 (2004), 597–620.

[7] N. Chomsky. 1965. *Aspects of the Theory of Syntax.* The MIT Press, Cambridge.

[8] H. H. Clark and S. E. Haviland. 1977. Comprehension and the given-new contract. *Discourse production and comprehension. Discourse processes: Advances in research and theory* 1 (1977), 1–40.

[9] M. Domínguez. 2017. *The Information Structure–Prosody Interface: On the Role of Hierarchical Thematicity in an Empirically-grounded Model.* Ph.D. Dissertation. Universitat Pompeu Fabra.

[10] M. Domínguez, M. Farrús, A. Burga, and L. Wanner. 2014. The Information Structure - Prosody Language Interface Revisited. In *Proceedings of the 7th International Conference on Speech Prosody*. Dublin, Ireland, 539–543.

[11] M. Domínguez, M. Farrús, A. Burga, and L. Wanner. 2016. Using hierarchical information structure for prosody prediction in content-to-speech applications. In *Proceedings of the 8th International Conference on Speech Prosody*. Boston, USA, 1019–1023.

[12] M. Domínguez, M. Farrús, and L. Wanner. 2016. Combining acoustic and linguistic features in phrase-oriented prosody prediction. In *Proceedings of the 8th International Conference on Speech Prosody*. Boston, USA, 796–800.

[13] M. Domínguez, M. Farrús, and L. Wanner. 2017. A Thematicity-based Prosody Enrichment Tool for CTS. In *Proceedings of the 18th Annual Conference of the International Speech Communication Association (INTERSPEECH 2017)*. Stockholm, Sweden, 3421–2.

[14] M. Domínguez, M. Farrús, and L. Wanner. 2018. Compilation of Corpora to Study the Information StructureâĂŞProsody Interface. In *11th edition of the Language Resources and Evaluation Conference (LREC2018)*. Mijazaki, Japan.

[15] M. Domínguez, M. Farrús, and L. Wanner. 2018. Thematicity-based Prosody Enrichment for Text-to-Speech Applications. In *9th International Conference on Speech Prosody 2018 (SP2018)*. Poznan, Poland.

[16] E Hajiĉova, B Partee, and P Sgall. 1998. *Topic-Focus Articulation, Tripartite Structures, and Semantic Content.* Kluwer Academic Publishers, Dordrecht.

[17] M.A.K. Halliday. 1967. Notes on Transitivity and Theme in English, Parts 1-3. *Journal of Linguistics* 3, 1 (1967), 37–81.

[18] Alfonso Igualada, Núria Estebe-Gibert, and Pilar Prieto. 2017. Beat gestures improve word recall in 3- to 5-year-old children. *Journal of Experimental Child Psychology* 156 (2017), 99–112.

[19] Frank Kügler, Bernadett Smolibocki, and Manfred Stede. 2012. Evaluation of Information Structure in Speech Synthesis : The Case of Product Recommender Systems Perception. In *ITG Conference on Speech Communication, IEEE*. 26–29.

[20] J Llanes-Coromina, I Vilà-Giménez, O Kushch, J. Borràs-Comes, and P. Prieto. 2018. Beat gestures help preschoolers recall and comprehend discourse information. *Journal of Experimental Child Psychology* 172 (2018), 168–188.

[21] I. A. Mel'čuk. 1988. *Dependency Syntax: Theory and Practice.* SUNY Press, Albany, NY. 400 pages.

[22] I. A. Mel'čuk. 2001. *Communicative Organization in Natural Language: The semantic-communicative structure of sentences.* Benjamins, Amsterdam, Philadephia. 393 pages.

[23] B. Mencía-López, D. Pardo, A. Trapote-Hernández, and L. A. Gómez-Hernández. 2013. Embodied Conversational Agents in Interactive Applications for Children with Special Educational Needs. In *Technologies for Inclusive Education: Beyond Traditional Integration Approaches*, David Griol Barres, Zoraida Callejas Carrión, and Ramón López-Cózar Delgado (Eds.). IGI Global, Hershey, USA, 59–88.

[24] D. Meurers, R. Ziai, N. Ott, and J. Kopp. 2011. Evaluating Answers to Reading Comprehension Questions in Context: Results for German and the Role of Information Structure. In *Proceedings of the TextInfer 2011 Workshop on Textual Entailment (TIWTE '11)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 1–9.

[25] A. Ortiz, M. del Puy Carretero, D. Oyarzun, J. J. Yanguas, C. Buiza, M. F. Gonzalez, and I. Etxeberria. 2007. *Elderly Users in Ambient Intelligence: Does an Avatar Improve the Interaction?* Springer Berlin Heidelberg, Berlin, Heidelberg, 99–114.

[26] D. Pérez-Marín and I. Pascual-Nieto. 2013. An exploratory study on how children interact with pedagogic conversational agents. *Behaviour & Information Technology* 32, 9 (2013), 955–964.

[27] M. Schröder and J. Trouvain. 2003. The German Text-to-Speech Synthesis System MARY: A Tool for Research, Development and Teaching. *International Journal of Speech Technology* 6, 4 (2003), 365–377. https://doi.org/10.1023/A:1025708916924

[28] R. Schwarzschild. 1999. GIVENness, AvoidF and Other Constraints on the Placement of Accent. *Natural Language Semantics* 7, 1 (1999), 141–177.

[29] K. Silverman, M. Beckman, J. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert, and J. Hirschberg. 2010. ToBI: A standard for labeling English prosody. In *Proceedings of Interspeech*. Makuhari, Japan, 146–149.

[30] P. Taylor and A. Isard. 1997. SSML: A Speech Synthesis Markup Language. *Speech Communication* 21, 1-2 (February 1997), 123–133.

[31] M. Vanrell, I Mascaró, F. Torres-Tamarit, and P. Prieto. 2013. Intonation as an Encoder of Speaker Certainty: Information and Confirmation Yes-No Questions in Catalan. *Language and Speech* 56, 2 (2013), 163–190. https://doi.org/10.1177/0023830912443942

[32] L. Wanner, E. André, J. Blat, S. Dasiopoulou, M. Farrús, T. Fraga, E. Kamateri, F. Lingenfelser, G. Llorach, O. Martínez, G. Meditskos, S. Mille, W. Minker, L. Pragst, D. Schiller, A. Stam, L. Stellingwerff, F. Sukno, B. Vieru, and S. Vrochidis. 2017. KRISTINA: A Knowledge-Based Virtual Conversation Agent. In *Proceedings of the 15th International Conference on Practical Applications of Agents and Multi-Agent Systems (PAAMS)*. Oporto, Portugal.

[33] L. Wanner, J. Blat, S. Dasiopoulou, M. Domínguez, G. Llorach, S. Mille, F. Sukno, E. Kamateri, S. Vrochidis, I. Kompatsiaris, et al. 2016. Towards a multimedia knowledge-based agent with social competence and human interaction capabilities. In *Proceedings of the 1st International Workshop on Multimedia Analysis and Retrieval for Multimodal Interaction*. ACM Digital Library, 21–26.

[34] P. Wargnier, G. Carletti, Y. Laurent-Corniquet, S. Benveniste, P. Jouvelot, and A. S. Rigaud. 2016. Field evaluation with cognitively-impaired older adults of attention management in the Embodied Conversational Agent Louise. In *2016 IEEE International Conference on Serious Games and Applications for Health, SeGAH 2016, Orlando, FL, USA, May 11-13, 2016*. 1–8.