*INTERSPEECH 2021: Show & Tell Contribution*
30 August – 3 September, 2021, Brno, Czechia

# ThemePro 2.0: Showcasing the Role of Thematic Progression in Engaging Human-Computer Interaction

*Mónica Domínguez*[1], *Juan Soler-Company*[1], *Leo Wanner*[1 2]

[1]Universitat Pompeu Fabra, Spain
[2]Institute for Catalan Advanced Research (ICREA), Spain
monica.dominguez@upf.edu, juan.soler@upf.edu, leo.wanner@upf.edu

## Abstract

Structuring speech into informative units is certainly a desirable feature in efficient human-machine communication. This paper introduces ThemePro 2.0, a toolkit that pre-processes long monologues into smaller cohesive units to be consumed by the text-to-speech module within a conversational agent. The methodology used is based upon the text's discourse structure modelled as thematic progression patterns. As shown in the demonstration, thematic progression modelling captures the underlying information structure at the discourse level and is, therefore, instrumental for cohesive speech output in the TTS component.

**Index Terms**: speech synthesis, text-to-speech, human-computer interaction, information structure, thematic progression

## 1. Introduction

Conversational agents often act as information providers in the context of question-answering interaction with humans. Recent advances in text-to-speech applications have achieved human-like voice quality that undoubtedly enhance this interaction. Yet, when the agent needs to speak for a long stretch of time uninterruptedly, information may not come across to the listener in an efficient and natural way that favours a good comprehension of the message.

Attention span of listeners is estimated to range between 30 to 60 seconds in human to human interaction [1]. Recent studies show that attention spans dramatically decrease in digital contexts to 8 to 12 seconds [2] and it has furthermore been proved that humans engage more with agents that show human-like expressiveness features (e.g. varied intonation, meaningful pauses, facial expressions, and body language) than with robotic interfaces [3]. However, how can humans engage with an agent that speaks uninterruptedly and gives no chance to users to take their turn in conversation? This may happen more often than expected in conversational agents acting as information providers when a long text (often retrieved from the web) is passed to the TTS module with the intention of providing users with the information they requested.

Presenting information into manageable segments to favor comprehension is instrumental for conversational agents in such scenario. If we are to engage our human listeners with an agent that presents human-like features, we cannot be content with any random generation of text chunks or rule-of-thumb approach to split long texts. Instead, we should explore the possibilities that discourse, information structure and recent advances in semantic representation have to offer to deal with such a task.

This paper introduces a demonstration of ThemePro 2.0; a web-based service for pre-processing monologues based on a thematic progression modelling strategy that guarantees cohesive units for a more meaningful spoken interaction. Section 2 explains how the analysis of thematic progression serves the purpose of monologue segmentation. Then, a use case scenario related to the H2020 project WELCOME[1] is introduced as a working example in Section 3. Section 4 presents the functionalities of ThemePro 2.0 and the main contribution of the toolkit. Finally, conclusions are briefly reported in Section 5.

## 2. Analysis of Thematic Progression

This section briefly motivates the reason why thematic progression serves the purpose of segmenting a long monologue into manageable units that support a better comprehension of the message.

Thematic progression departs from the theoretical framework of information structure (IS) (cf. among others [4, 5, 6]). IS states that a message is built based on previous or existing information (a theme). In other words, new information in a sentence (also known as the rheme[2]) is added to known information (i.e., the theme): a rheme says something about a theme. Such a structure, which is incrementally more complex beyond sentence level (i.e. at the discourse level), is described by thematic progression patterns.

Studies show that information structure is often matched (especially by efficient communicators) to a varied expressive prosody, where pauses (among other prosodic features) based on IS play a central role [8]. Moreover, pauses allow listeners to take time and process information efficiently and let them take their turn in human-computer interaction.

### 2.1. Main contribution of ThemePro 2.0

A preliminary implementation for the analysis of thematic progression was introduced in [9]. The work presented in this paper revisits the previous implementation and builds a new functionality to adapt communicative units to monologues fed to the TTS engine.

The main functionality of the original implementation of ThemePro [9] was the visualization capabilities of the toolkit and the analysis of thematic progression of texts. The main contribution of our work departs on this analysis and expands it with an algorithm to segment monologues into smaller cohesive units in the context of information provided by a conversational agent. We are furthermore deploying a TTS service to test the spoken output.

[1]WELCOME provides a holistic platform to assist in the reception, orientation and integration of migrant and refugee communities to host countries.

[2]Different authors refer to theme and rheme using different terminology (see [7] for a full reference on IS and its interfaces).

Copyright © 2021 ISCA

### 2.2. Natural Language Processing and Speech Technologies in ThemePro 2.0

ThemePro 2.0 includes several state-of-the-art Natural Language Processing (NLP) technologies, such as syntactic parsing [10], thematicity parsing [11], word embeddings (Word2vec Google News' word embeddings[3]) and co-reference resolution (Neuralcoref by Hugging Face[4]). As TTS service, we have deployed Mozilla TTS[5].

## 3. Use Case Description and Examples

We envisage ThemePro 2.0 as component of a conversational agent that provides information prompted by voice interaction upon request of the user in the context of the WELCOME H2020 European project[6]. The WELCOME project involves user partners from Catalonia (Spain), Germany and and Greece that specify relevant functionalities for their migrant and refugee communities. Users of the WELCOME platform can get relevant information about services provided by local authorities (among other functionalities).

## 4. ThemePro 2.0 Segmentation Algorithm

The novel feature of ThemePro 2.0 is the segmentation algorithm that is built upon the theoretical framework of thematic progression. In our implementation, a cohesive thematic progression is considered within a pre-determined length restriction. Such length has been empirically derived after testing the Mozilla TTS application that already handles the problem of limited input length in neural architectures with a double decoder consistency (DDC) architecture [12]. The algorithm adjusts the length of the text sent to the TTS engine considering the established constraints and the most cohesive thematic progression pattern. Cohesiveness is assessed on theme spans ($T_n$) computing the cosine similarity based on the span's centroid (which is computed as the average of all word embeddings that compose the span) from $T_n$ to both $T_{n-1}$ and $R_{n-1}$. Whenever two consecutive spans have a similarity that is higher than an empirically-determined threshold, we add sentence n to the same block as sentence n-1.

Co-reference resolution is also considered for those cases where semantic similarity based on word embeddings cannot possibly perform well, that is mostly, personal pronouns. Whenever co-referent terms appear in both $T_n$ and $T_{n-1}$ or $R_{n-1}$, we add sentence n to the same block as sentence n-1. The code is made available as a Docker image in the following repository: `https://github.com/monikaUPF/ThemePro_2.0`

## 5. Conclusions

This paper introduces ThemePro 2.0, a toolkit designed for the pre-processing of long monologues into comprehensive communicative units based on thematic progression. The main use case of such a module is in a conversational agent architecture, right before the TTS application. The main advantages of using communicatively cohesive units are: (1) the listener comprehension is enhanced, (2) dialogue turns can be taken by the listener between meaningful pauses between these units.

A communicatively-motivated discourse segmentation is an essential corner stone to advance on more complex interactive strategies, such as prosodic cues, avatar movement, facial expressions and so on. Implementing prosodic and visual expressiveness upon a sound discourse structure is the underlying meaningful layer of reference to achieve fully efficient and user-oriented conversational agents.

[3]Word2vec: https://code.google.com/archive/p/word2vec/

[4]Neuralcoref: https://github.com/huggingface/neuralcoref

[5]MozillaTTS: https://github.com/synesthesiam/docker-mozillatts

[6]See more information on the project as well as other functionalities of the WELCOME platform like virtual reality and visual analytics in https://welcome-h2020.eu.

## 6. Acknowledgements

This work has received funding from the European Commission under Grant n. H2020-870930.

## 7. References

[1] H. Jin, L. Fang, R. Wang, X. Li, Y. Zheng, and Y. Yang], "Prosodic boundaries in speech: A window to spoken language comprehension," *Advances in Psychological Science*, vol. 29, no. 3, pp. 425–437, 2021.

[2] N. Geri, A. Winer, and B. Zaks, "Challenging the six-minute myth of online video lectures: Can interactivity expand the attention span of learners?" *The Online Journal of Applied Knowledge Management (OJAKM)*, 2017.

[3] S. Irshad and A. Perkis, "Increasing user engagement in virtual reality:the role of interactive digital narratives to trigger emotional responses," in *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, ser. NordiCHI '20. New York, NY, USA: Association for Computing Machinery, 2020. [Online]. Available: https://doi.org/10.1145/3419249.3421246

[4] F. Daneš, "One instance of Prague School methodology: Functional analysis of utterance and text," *Garvin*, pp. 132–141, 1970.

[5] ——, "Functional sentence perspective and the organization of the text." in *Papers on Functional Sentence Perspective*. The Hague: Mouton, 1974, pp. 106–128.

[6] E. Hajičová and J. Mírovský, "Discourse Coherence Through the Lens of an Annotated Text Corpus: A Case Study," in *Proceedings the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, 2018, pp. 1637–1642.

[7] L. Mereu, Ed., *Information Structure and its Interfaces*. Berlin, Boston: De Gruyer Mouton, 2009.

[8] M. Domínguez, M. Farrús, and L. Wanner, "The information structure–prosody interface in text-to-speech technologies. an empirical perspective," *Corpus Linguistics and Linguistic Theory*, 2021.

[9] M. Domínguez, J. Soler, and L. Wanner, "ThemePro: A toolkit for the analysis of thematic progression," in *Proceedings of the 12th Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, 2020, pp. 1000–1007.

[10] M. Honnibal, I. Montani, S. Van Landeghem, and A. Boyd, "spaCy: Industrial-strength Natural Language Processing in Python," 2020. [Online]. Available: https://doi.org/10.5281/zenodo.1212303

[11] M. Domínguez Bajo, A. Burga, M. Farrús, and L. Wanner, "Towards expressive prosody generation in tts for reading aloud applications," *IberSpeech 2018; 2018 Nov 21-23; Barcelona, Spain. Baixas, France: ISCA; 2018. p. 40-4.*, 2018.

[12] G. E., "Solving attention problems of tts models with double decoder consistency," 2020. [Online]. Available: erogol.com/solving-attention-problems-of-tts-models-with-double-decoder-consistency/