Detecting Discourse Creativity in Chat Conversations

Costin-Gabriel Chiru and Traian Rebedea

University Politehnica of Bucharest, Department of Computer Science and Engineering,
313 Splaiul Independetei, Bucharest, Romania
{costin.chiru,traian.rebedea}@cs.pub.ro

Abstract. In this paper we propose a new method for identifying creativity that is based on analyzing a corpus of chat conversations on the same topic and extracting the new ideas expressed by participants. The application is a first step in supporting creativity in online group discussions by highlighting the novel concepts present in conversations (*new ideas*) and also by identifying topics that could have become important, if not forgotten during the debates (*lost ideas*).

Keywords: Creativity, Linguistics, CSCL, Learning Analytics.

1 Introduction

Creativity is defined as the ability to transcend traditional ideas, patterns, relationships into meaningful new ideas, interpretations, etc. [4]. In written discourse, it is called linguistic creativity [3] and it measures "new and creative ways of expressing a given idea". Linguistic creativity is sometimes described as an unpredictable departure from the rules of regular word formation [2]. In our opinion, linguistic creativity could be defined as a deviation from the standard way of writing on different topics.

We propose a new approach for assessing discourse creativity based on processing a large volume of chat conversations on the same topic. Thus, we see the rare concepts (that are not off-topic) as being intentional deviations from the standard concepts that are debated in the corpus of chats, representing a sign of discourse and linguistic creativity. This way we link creativity to *differences in the participants' discourses*.

2 Corpora and Application

The corpora used for analysis consists of chats of students that worked in small teams to find the best online instrument for information sharing in a company [1]. We analyzed the content of 62 chat conversations in order to detect significant similarities and differences between them. We considered that the similarities are represented by common concepts that were uttered by the majority of the participants, while the differences – concepts present only in very few conversations – should be signs either of creativity or divergent thinking or of off-topic content. In order to discriminate between these two opposite situations, we evaluated the utterances to detect which ones were on topic. Therefore, we considered that the on-topic words were the ones from the lexical field of concepts such as "blog", "chat", "forum", "wiki", or "web".

2.1   Lost Ideas, Similar Ideas, Ideas and Reactions

The *lost ideas* represent the concepts found only in a small number of chats. For their detection, we first identified the rare words from the corpus (the ones found only once) and afterwards checked to see if they were on-topic by evaluating the utterances containing them (whether they also contained on-topic concepts or not). *New ideas/similar ideas* express concepts that have a high frequency in a small number of chats. For each utterance where we found such a concept, we extracted a text pattern (a window of type "*word₁ [*] word₂*") and used this pattern to detect similar situations in the other chats.

We have also devised a method that classifies the participants to chats into 5 classes according to their discourse creativity: leaders, developers, innovators, negativists, and conclusives. This list is open-ended so there may also be other typologies. In order to detect the typology of a participant, it is first necessary to automatically classify the chat utterances according to their content in the following 5 different classes:

— *Ignored utterances* – usually they contain social noise.
— *Ideas* – utterances containing concepts that are on-topic.
— *Conclusions* – contains patterns that are specific to conclusive utterances.
— *Developments to conclusions* – if the previous utterance contains some on-topic concepts, its development should contain words from the same specter.
— *Reactions* – approvals, disapprovals or continuations of previously stated ideas.

3   Conclusions

Discourse creativity is a concept that is very difficult to assess automatically. In this paper, we have presented a method for determining creativity in a large collection of online discussions that debate the same topic within a CSCL scenario. Moreover, the proposed method can be used to identify typologies of participants starting from the elements used for assessing creativity.

References

1. Rebedea, T., Dascalu, M., Trausan-Matu, S., Armitt, G., Chiru, C.: Automatic Assessment of Collaborative Chat Conversations with PolyCAFe. In: Kloos, C.D., Gillet, D., Crespo García, R.M., Wild, F., Wolpers, M. (eds.) EC-TEL 2011. LNCS, vol. 6964, pp. 299–312. Springer, Heidelberg (2011)
2. Renouf, A.: Tracing lexical productivity and creativity in the british media: the chavs and the chav-nots. In: Lexical Creativity, Texts and Contexts, pp. 61–89. John Benjamins Publishing Company, Amsterdam (2007)
3. Veale, T.: Creative language retrieval: A robust hybrid of information retrieval and linguistic creativity. In: Proceedings of ACL 2011, Portland, Oregon, USA, pp. 278–287 (2011)
4. Zhu, X., Xu, Z., Khot, T.: How creative is your writing? a linguistic creativity measure from computer science and cognitive psychology perspectives. In: Proceedings of the Workshop on Computational Approaches to Linguistic Creativity, pp. 87–93. ACL (2009)