Discourse cohesion, a signature of collaboration

Mihai Dascalu, Stefan Trausan-Matu, Philippe Dessus, Danielle S. Mcnamara

Discourse cohesion: A signature of collaboration

Mihai Dascalu,
Stefan Trausan-Matu
University Politehnica of Bucharest
313 Splaiul Indepententei
Bucharest, Romania
mihai.dascalu@cs.pub.ro
stefan.trausan@cs.pub.ro

Philippe Dessus
Univ. Grenoble Alpes, LSE,
F38000 Grenoble, France
philippe.dessus@upmf-grenoble.fr

Danielle S. McNamara
Arizona State University
Tempe, Arizona 85287-2111
dsmcnama@asu.edu

ABSTRACT

As Computer Supported Collaborative Learning (CSCL) becomes increasingly adopted as an alternative to classic educational scenarios, we face an increasing need for automatic tools designed to support tutors in the time consuming process of analyzing conversations and interactions among students. Therefore, building upon a cohesion-based model of the discourse, we have validated *ReaderBench*, a system capable of evaluating collaboration based on a social knowledge-building perspective. Through the inter-twining of different participants' points of view, collaboration emerges and this process is reflected in the identified cohesive links between different speakers. Overall, the current experiments indicate that textual cohesion successfully detects collaboration between participants as ideas are shared and exchanged within an ongoing conversation.

Categories and Subject Descriptors

I.2.7 [**Natural Language Processing**], K.3.1 [**Computer Uses in Education**]

General Terms

Algorithms, Measurement

Keywords

Computer Supported Collaborative Learning, Cohesion, Discourse Analysis, Learning Analytics, Collaboration Assessment.

1. INTRODUCTION

Computer Supported Collaborative Learning (CSCL) has gained a broader usage in multiple educational scenarios and has become a viable alternative to classic learning environments and settings as it can be employed in a multitude of activities, such as MOOCs or collaborative serious games. At the same time, with its wider adoption, the need for automated tools capable of supporting and evaluating the corresponding actors has becomes more stringent due to the fact that the analysis of conversations involving multiple participants is a time consuming process. Trausan-Matu [23], for example, reported that the time required for a thorough analysis of a chat session greatly exceeds the actual duration of the online conversation, rendering the manual evaluation process virtually impossible for large corpora of conversations.

In a nutshell, collaboration can be perceived as a measure of interaction among participants centered on sharing ideas, fostering creativity for working in groups [24] and influencing others' points of view during the discussion. Therefore, our interest consists of automatically assessing collaboration from CSCL text-based interactions among multiple participants performed during specific educational scenarios.

From a more pragmatic perspective, our aim is to validate a computational model for evaluating collaboration based on a cohesion-based model of the discourse [6; 25]. This study represents an extension of the initial model [6], which has now been further validated within an educational setting. Therefore, besides fine-tuning the automatic assessment process, validation has now been performed on a larger corpus manually annotated by considerably more evaluators. As an overview, we perform a longitudinal analysis of an ongoing conversation, following its timeline and relying on cohesion to model the knowledge transferred or constructed among participants. In other words, CSCL outlines a socio-cultural paradigm focused on the idea that new knowledge is created collaboratively through the process of social knowledge building [2; 20; 22].

In contrast to the most simplistic models of assessing collaboration which rely on counting the number of utterances exchanged between or addressed to different speakers, our model supports the notion that cohesion is a salient predictor of collaboration. Therefore, by modeling the interactions between participants through textual cohesion, signatures of collaboration emerge. It is common for tutors to attempt to detect breaks in conversations that have limited or no collaboration or intense collaboration zones in learners' productions. Automatic methods, such as *ReaderBench* [5], will provide crucial support to tutors in extracting such zones.

The following section is centered on the concept of cohesion, as well as underlying computational approaches for analyzing students' productions. The third section briefly introduces the mechanism of scoring utterances or contributions, while the fourth section presents the collaboration assessment model. Then we shift the point of interest towards the validation of the proposed model, followed by discussions and conclusions.

2. DISCOURSE COHESION

The concept of *cohesion* was introduced by Halliday and Hasan [9] in terms of the cohesive ties perceived as "relations of meaning that exist within the text, and that define it as a text." [9]. Cohesion provides overall unity and is used for establishing the underlying structure of meaning. Therefore, cohesion addresses the connections in a text based on features that highlight relations between constituent elements (words, sentences or blocks of text). In other words, text cohesion can be perceived as the sum of lexical, grammatical, and semantic relations that link together textual units.

From a computational viewpoint, cohesion is reflected in the linguistic form of discourse [16] and is often regarded as an

indicator of its structure. More specifically, cohesion can derive from: a) discourse connectedness reflected as relations between sentences (e.g., explanation, contrast) through cue words or phrases (e.g., but, because) [15]; b) referencing expressions that reflect the status of an entity in the discourse and can be identified through co-reference resolution [11; 15; 18]; c) lexical or semantic similarity of words obtained from semantic distances in ontologies [4], cosine similarity applied on vector spaces from Latent Semantic Analysis [13], or topic relatedness measured as Jensen-Shannon divergence in Latent Dirichlet Allocation [3].

Within our implemented model, cohesion is determined as the product between the *inverse normalized distance between textual elements* and an average semantic similarity measure composed of: a) *lexical proximity* that is easily identifiable through identical lemmas and semantic distances [4] within ontologies (*WordNet* [17] or, for French language, a transposed and serialized version of *Wordnet Libre du Français* (WOLF) [19]); and b) *semantic similarity* measured through Latent Semantic Analysis (LSA) [13] and Latent Dirichlet Allocation (LDA) [3]. Additionally, specific natural language processing techniques [14] are applied to reduce noise and to improve the system's accuracy: spell-checking, tokenization, splitting, part of speech tagging, parsing, stop words elimination, dictionary-only words selection, stemming, lemmatization, named entity recognition, and co-reference resolution [18].

LSA and LDA models were trained using three specific corpora: *TextEnfants* [7] (approx. 4.2M words), *Le Monde* (French newspaper, approx. 24M words) for French, and Touchstone Applied Science Associates (TASA) corpus (approx. 13M words) for English. Moreover, improvements have been enforced on the initial models: the reduction of inflected forms to their lemmas, the annotation of each word with its corresponding part of speech through a NLP processing pipe, and the adjustment of occurrences through the use of term frequency-inverse document frequency (*Tf-Idf*) [14].

Our previous experiments [5] have shown that Wu-Palmer ontology-based semantic similarity [27] combined with LSA and LDA models can be used to complement each other, in the sense that underlying semantic relationships are more likely to be identified if multiple complementary approaches are combined after normalization, reducing the errors that can be induced by using a single method. Overall, in order to better evaluate cohesion between textual fragments, we have combined information retrieval specific techniques, mostly reflected in word repetitions and normalized number of occurrences, with semantic distances extracted from ontologies or from LSA-based or LDA-based semantic models. Based on the previous cohesion function, we define a cohesive link as a connection between two textual elements that has a high value for cohesion. In the actual implementation, the mean value of all semantic similarities is considered the threshold for detecting cohesive links.

In the end, in order to have a better representation of discourse in terms of underlying cohesive links, we represent these links as a cohesion graph [6; 25], which can be perceived as a generalization of the previously proposed utterance graph [26].

3. UTTERANCE SCORING

Assessing collaboration proves to be challenging, particularly when our aim consists of finding an approximation that best reflects the importance or the value of a collaborative act. Nevertheless, the evaluation process must be performed in two steps. First, the local importance of each utterance must be determined with regards to the entire conversation. Second, collaboration is approximated as the impact of each utterance on another participant's discourse through cohesive links. This section is centered on evaluating each utterance and assigning it an importance score, whereas the next section details the process of automatically assessing collaboration.

In order to evaluate the importance of each utterance, we must first determine the value of its constituents or, more specifically, the relevance of each contained word. With regards to the process of evaluating each word's relevance in relation to its corresponding textual fragment (e.g., sentence, utterance, or entire conversation), there are several classes of factors that play an important role in the final analysis (see Table 1).

Table 1. Factors used to measure a word's relevance

Class	Descriptors
Statistical presence	Normalized term frequency used to reflect the specificity of each single text
Semantic relatedness	Semantic similarity to the analysis element (sentence, utterance, entire conversation)
Semantic coverage	The importance of the semantic chain containing a particular word and its span throughout the entire conversation

Out of the three classes, the most straightforward factor consists of computing the statistical presence of each word. The next class is focused on determining the semantic relatedness between a word and its corresponding textual fragment, whereas the last class evaluates the semantic coverage of each concept. Semantically related words are grouped together in lexical chains by using an adaptations of the algorithm proposed by Galley and McKeown [8] that relies solely on semantic distances from WordNet [4]. The previous lexical chains are then merged into semantic chains by using the semantic similarities between the chains expressed in terms of LSA and LDA models [5]. Semantic coverage is reflected in the length and the span of these semantic chains as this provides a reliable global estimator for the importance of each concept with regards to the entire conversation. Based on the previous classes of factors, the keywords of the conversation are determined as the words with the highest cumulative relevance based on their individual occurrences.

In terms of the scoring model, each utterance is initially assigned an individual score equal to the normalized term frequency of each word multiplied by its previously determined relevance [5]. Putting it differently, we measure to what extent each utterance conveys the main concepts of the overall conversation, as an estimation of on-topic relevance. Afterwards, these individual scores are augmented through cohesive links to other inter-linked textual elements by using the previously defined cohesion values as weights. Overall, we can state that keywords are used to reflect the local importance of each word, whereas cohesive links are used to transpose the local relevance upon other inter-linked elements.

4. AUTOMATIC COLLABORATION ASSESSMENT

The actual information transfer through cohesive links from the cohesion graph can be split into a personal and a social knowledge-building (KB) process [2; 20; 22] taking place within

each participant's utterances. First, a personal dimension emerges by considering utterances from the same speaker as a continuation of one's discourse. Second, utterances exchanged with different speakers encompass collaboration and sustain social interaction. Therefore, each utterance now has its previously defined importance score and a knowledge-building effect, both personal and social. By considering all cohesive links, the values for personal and social knowledge building are correspondingly increased: if the link is between utterances having the same speaker, the previous KB index (sum of personal and social KB) is transferred to the personal dimension of the current utterance; otherwise, if the pair of utterances is between different participants, the social knowledge-building dimension of the currently analyzed utterance is increased with the same amount of information (KB index multiplied by the cohesion function). In other words, continuation of ideas or explicitly referencing utterances of the same speaker builds a personal knowledge-building effect, whereas the social perspective measures the interaction with other participants.

Within the conducted experiments, collaboration emerges from: a) the active participation of all members, b) the sharing of ideas and points of view, c) the dense inter-twining of utterances whose reference IDs are manually annotated by the speakers through an explicit referencing facility available from the conversation environment – ConcertChat [10]

Our automated text analysis tool, *ReaderBench* [5], highlights the contribution of utterances through new layers of granularity added to the overall collaboration analysis. By new layers, we refer to the estimation of both personal and social knowledge-building effects, as well as the identification of intense collaboration zones – intervals of utterances in which participants are actively involved, collaborate, and generate new ideas related to the ongoing context of the discussion. From a computational perspective, we used a greedy algorithm [6] in order to build up intense collaboration zones by expanding maximum local values or social knowledge-building peaks. Each local maximum is expanded sideways and, in the end, only zones above a minimum spread of 5 utterances are selected as intense collaboration zones.

5. VALIDATION OF COLLABORATION ASSESSMENT

The validation experiments focused on the assessment of 10 chat conversations that took place in an academic environment in which Computer Science students from the 4th year undergoing a Human-Computer Interaction course in Romania debated on the advantages and disadvantages of CSCL technologies. According to the script presented in the previous section, each conversation involved 4 to 5 participants who each had to argue for a given technology (e.g., chat, blog, wiki, forum, or Google Wave) in specific usage scenarios during the first phase of the discussion, and then subsequently propose an integrated alternative that encompassed the previously presented advantages. The 10 chats were manually selected from a larger corpus of over 100 conversations. The chat conversations were manually annotated by 76 4th year undergraduate students following the same course, but from a different class, and 34 freshman master students attending the Adaptive and Collaborative Systems course. Each student annotated 3 chat conversations while addressing the following tasks: a) grading each speaker on a 1 to10 scale in terms of collaboration or exchange of ideas with other participants, and

b) identifying intense collaboration zones as segments of the conversations in which multiple participants contribute and actively participate in the ongoing discourse with on-topic and relevant utterances.

A significant amount of time is necessary for a rater to glean a deep understanding of a conversation (1.5 to 4 hours on average). Hence, we opted to distribute the evaluation of each conversation to multiple raters [23]. This approach resulted in an average of 33 annotations per conversation.

First, we validated the machine versus human agreement by computing intra-class correlations between raters for each chat and, second, as these correlations were all very high indicating very few disagreements between raters, non-parametric correlations (Spearman's Rho) were calculated between machine and human mean ratings for each chat (see Table 2). The low number of participants per chat was also a determinant factor for the previous high values, as the comparisons between shorter numerical series tend to have more extreme values, resulting in either highly positive or negative correlations.

As an interpretation of the results presented in Table 2, we can observe that the non-parametric correlations were high except for chats 3 and 8. In the latter conversations, we were able to identify atypical behaviors that justify these discrepancies: a) dominance of the conversation by certain participants at given moments throughout the conversation; b) existence of wide-spread segments in which multiple participants seemed to get involved in a similar degree, rendering the differentiation among them more difficult; c) disequilibrium of the conversation due to the focus on only one technology ("blog" in both conversations) which shifted the overall balance with regards to the other technologies that should have been debated.

With regards to the identification of intense collaboration zones, all manual annotations were cumulated within a histogram, which presented for each utterance the number of raters that considered it to be part of an intense collaboration. Afterwards, the same greedy algorithm was applied on this histogram in order to obtain an aggregated version that reflected the intense collaboration zones emerging as an overlap of all annotations (see Table 2).

Moreover, as presented in Table 2, there is a good overlap in terms of accuracy (measured as precision, recall and F1 score) between the annotated collaboration zones and those that were automatically identified. This indicates that the model is a good estimator of the annotated zones.

However, the rather low correlations with an average of .34 in terms of collaboration evolution are justifiable as the scales are completely different. On the one hand, we have the number of inclusions of each utterance within manually identified intense collaboration zones. This was a subjective and bias-prone task as there were no constraints imposed in terms of the overall coverage of these zones and the rater's perception of interaction among multiple participants. Additionally, this is a quantitatively cumulative score obtained from the overlap of zones from multiple raters which, in essence, is a transversal sum of occurrences. On the other hand, the system provides an estimation of social knowledge building based on the social interaction modeled through cohesive links. The latter estimation reflects a qualitative process following the timeline spanning the conversation and therefore a longitudinal effect in contrast to the previous transversal sum.

Table 2. Intense collaboration zones extracted from manual versus automatically identified annotations

Conver-sation	No. utter-ances	ICC	Rho	Zones extracted from manual annotations	Zones automatically identified	P	R	F1 score	r
Chat 1	339	.97	1.0**	[37;128], [167;298]	[16;130], [154;228], [238;279], [291;298], [307;321], [334;345]	.76	.91	.83	.27
Chat 2	283	.82	.80	[13;49], [68;131], [144;169], [193;207], [229;236], [245;266]	[10;47], [55;79], [88;112], [128;237], [246;275]	.64	.85	.73	.31
Chat 3	405	.73	.30	[36;315]	[15;353], [397;401], [364;373]	.79	1.0	.89	.43
Chat 4	251	.91	.70	[19;148], [184;194], [203;212]	[27;131], [188;209], [220;259]	.73	.79	.76	.44
Chat 5	416	.96	.90*	[28;98], [120;265], [280;287], [346;362]	[16;285], [308;314], [331;407]	.68	.99	.81	.23
Chat 6	378	.96	.98**	[64;227], [248;308], [321;359]	[47;331], [344;371]	.80	.95	.87	.46
Chat 7	270	.91	.70	[40;97], [108;128], [145;220], [232;257]	[14;126], [137;213], [221;236], [255;265]	.73	.85	.78	.21
Chat 8	389	.92	.40	[30;127], [140;154], [189;208], [235;285], [299;356]	[14;43], [59;172], [193;208], [238;250], [260;384]	.71	.87	.78	.31
Chat 9	190	.97	.80	[51;65], [80;190]	[48;67], [84;121], [127;184], [190;195]	.91	.86	.89	.56
Chat 10	297	.86	.80	[27;75], [89;104], [139;287]	[18;152], [167;182], [190;221], [234;243], [253;257], [267;305]	.69	.76	.72	.21
Average	*321.8*	*.90*	*.74*			*.74*	*.88*	*.81*	*.34*

Note: *p < .05; **p < .01

6. Discussion

Based on the results presented in Table 2 and highly related to the process of modeling social knowledge building, we can consider cohesion as a binder between the utterances within an intense collaboration zone. Cohesion measures the topic relatedness between the utterances, whereas social interaction in a cohesive context determines collaboration. In other words, cohesion among utterances of different speakers becomes a signature of collaboration.

Nevertheless, we must highlight certain limitations of our model. Foremost, the model addresses only specific educational situations in which participants share, continue, debate, or argue certain topics or key concepts of the conversation. In other words, collaboration is particularly derived from idea sharing between participants who exchange cohesive utterances. It becomes obvious that specific discourse markers or speech acts (e.g., confirmations or negations) [1; 21] should be also considered for modeling collaboration, but for our specific educational scenario presented in detail in section five, cohesion by itself proved to be a reliable predictor. As the students debated on specific topics, textual cohesion highlighting the exchange or continuation of ideas represented a reliable estimator of the generated collaborative effect.

Overall, the presented model should not be perceived as a rigid structure, but as an adaptable one that evolves based on the cohesion to other participants' utterances. Nevertheless, we must highlight additional limitations in terms of personal knowledge building, social knowledge transfer, actual noise of the experiment, and underlying cognitive processes. As an initial assumption, we consider personal knowledge building as the reflection of one's thoughts continued into subsequent utterances

through cohesive links. This is only partially true because the underlying cognitive processes can be more elaborate than the written form expressed within the conversation.

7. CONCLUSIONS

Starting from a dialogic model of discourse represented through cohesive links, we validated our system in terms of analyzing collaboration by employing an assessment model based on social knowledge building. This demonstrated that the microstructure level connectedness reflected in cohesion is a building block for achieving a truly collaborative discourse.

Overall, collaboration was assessed using a bottom-up approach. Initially, the importance of an utterance was measured with regards to the overall discourse, it was assigned a corresponding score, and afterwards the impact on other utterances of different speakers was determined. In other words, within each intense collaboration zone, there are multiple utterances that are cohesive one to another and whose local importance scores are used to the approximate the collaborative effect. This approach involves a twofold estimation that uses an approximation of each utterance's importance and considers that the transferred information between different participants as a measure for collaboration.

Based on this analysis, we can extend the perspective of collaboration in terms of achieving a coherent representation of the discourse through the inter-animation of participants' points of view. Therefore, starting from dialogism as a framework of CSCL [12], our aim is to employ methods specific to computational linguistics in order to model the exchange and sharing of ideas among participants in a conversation.

Importantly, our analyses have a broad spectrum of applications, extending from the initial evaluation based on cohesion between

utterances towards group cohesion achieved through collaboration.

8. ACKNOWLEDGEMENTS

9. REFERENCES

[1] Austin, J.L., 1962. *How to Do Things With Words*. Harvard University Press, Cambridge, MA.

[2] Bereiter, C., 2002. *Education and mind in the knowledge age*. Lawrence Erlbaum Associates, Mahwah, NJ.

[3] Blei, D.M., Ng, A.Y., and Jordan, M.I., 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research 3*, 4-5, 993–1022.

[4] Budanitsky, A. and Hirst, G., 2006. Evaluating WordNet-based Measures of Lexical Semantic Relatedness. *Computational Linguistics 32*, 1, 13–47.

[5] Dascalu, M., 2014. *Analyzing discourse and text complexity for learning and collaborating, Studies in Computational Intelligence*. Springer, Switzerland.

[6] Dascalu, M., Trausan-Matu, S., and Dessus, P., 2013. Cohesion-based analysis of CSCL conversations: Holistic and individual perspectives. In *CSCL 2013*, N. Rummel, M. Kapur, M. Nathan and S. Puntambekar Eds. ISLS, Madison, USA, 145–152.

[7] Denhière, G., Lemaire, B., Bellissens, C., and Jhean-Larose, S., 2007. A semantic space for modeling children's semantic memory. In *Handbook of Latent Semantic Analysis*, T.K. Landauer, D.S. McNamara, S. Dennis and W. Kintsch Eds. Erlbaum, Mahwah, 143–165.

[8] Galley, M. and Mckeown, K., 2003. Improving word sense disambiguation in lexical chaining. In *IJCAI'03*, G. Gottlob and T. Walsh Eds. Morgan Kaufmann Publishers, Inc., Acapulco, Mexico, 1486–1488.

[9] Halliday, M.a.K. and Hasan, R., 1976. *Cohesion In English*. Longman, London.

[10] Holmer, T., Kienle, A., and Wessner, M., 2006. Explicit Referencing in Learning Chats: Needs and Acceptance. In *EC-TEL 2006*, W. Nejdl and K. Tochtermann Eds. Springer, Crete, Greece, 170– 184.

[11] Jurafsky, D. and Martin, J.H., 2009. *An introduction to Natural Language Processing. Computational linguistics, and speech recognition*. Pearson Prentice Hall, London.

[12] Koschmann, T., 1999. Toward a dialogic theory of learning: Bakhtin's contribution to understanding learning in settings of collaboration. In *CSCL'99*, C.M. Hoadley and J. Roschelle Eds. ISLS, Palo Alto, 308–313.

[13] Landauer, T.K. and Dumais, S.T., 1997. A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychological Review 104*, 2, 211–240.

[14] Manning, C.D. and Schütze, H., 1999. *Foundations of statistical Natural Language Processing*. MIT Press, Cambridge, MA.

[15] McNamara, D.S., Graesser, A.C., Mccarthy, P., and Cai, Z., 2014. *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press, Cambridge.

[16] McNamara, D.S., Louwerse, M.M., Mccarthy, P.M., and Graesser, A.C., 2010. Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes 47*, 4, 292–330.

[17] Miller, G.A., 1995. WordNet: A lexical database for English. *Communications of the ACM 38*, 11, 39–41.

[18] Raghunathan, K., Lee, H., Rangarajan, S., Chambers, N., Surdeanu, M., Jurafsky, D., and Manning, C.D., 2010. A Multi-Pass Sieve for Coreference Resolution. In *EMNLP '10* ACL, Cambridge, MA, 492–501.

[19] Sagot, B. and Darja, F., 2008. Building a free French wordnet from multilingual resources. In *Ontolex 2008*, Marrakech, Maroc, 6.

[20] Scardamalia, M., 2002. Collective cognitive responsibility for the advancement of knowledge. In *Liberal Education in a Knowledge Society*, B. Smith and C. Bereiter Eds. Open Court Publishing, Chicago, 67–98.

[21] Searle, J., 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, Cambridge, UK.

[22] Stahl, G., 2006. *Group cognition. Computer support for building collaborative knowledge*. MIT Press, Cambridge, MA.

[23] Trausan-Matu, S., 2010. Automatic Support for the Analysis of Online Collaborative Learning Chat Conversations. In *3rd Int. Conf. on Hybrid Learning*, P.M. Tsang, S.K.S. Cheung, V.S.K. Lee and R. Huang Eds. Springer, Beijing, China, 383–394.

[24] Trausan-Matu, S., 2010. Computer support for creativity in small groups using chats. *Annals of the Academy of Romanian Scientists, Series on Science and Technology of Information 3*, 2, 81–90.

[25] Trausan-Matu, S., Dascalu, M., and Dessus, P., 2012. Textual complexity and discourse structure in Computer-Supported Collaborative Learning. In *ITS 2012*, S.A. Cerri, W.J. Clancey, G. Papadourakis and K. Panourgia Eds. Springer, Chania, Grece, 352–357.

[26] Trausan-Matu, S., Stahl, G., and Sarmiento, J., 2007. Supporting polyphonic collaborative learning. *Indiana University Press, E-service Journal 6*, 1, 58–74.

[27] Wu, Z. and Palmer, M., 1994. Verb semantics and lexical selection. In *ACL '94* ACL, New Mexico, USA, 133–138.