Leveraging Natural Language Processing to Uncover Comprehension Processes in Constructed Responses

Laura K. Allen,[a] Lauren E. Flynn,[a] & Danielle S. McNamara [b]
University of Minnesota,[a] Arizona State University [b]

Natural Language Processing (NLP) is a way to analyze language with computers; NLP techniques can model the language of students' discourse across multiple dimensions[1-2]

what do we use as input?

- student essays
- learning reflections
- think-alouds
- open-ended reading comprehension questions

what do we get as output?

insight into students':

- emotions/affect
- cognitive processes
- source use in essays
- writing sophistication

The Multiple Dimensions of NLP

word level

Red blood cells are a necessity for the body. They bring carbon dioxide to the cells of the body. The body then turns the oxygen into carbon dioxide. The red blood cells take the carbon dioxide and have it removed

……

healthy red blood cells carry oxygen around your body. Sometimes not enough oxygen is transported.

sentence level

Not enough oxygen results in anemia.

……

paragraph level

Anemia is a condition where not enough oxygen gets into the body. Anemia can make a person feel tired and weak. One time, my doctor told me I was anemic and it made me feel really tired. I guess, anemia must have something to do with problems with your red blood cells. Is blood a cell?

document level

lexical :

- academic word use
- abstract/concrete word use
- word imageability

semantic :

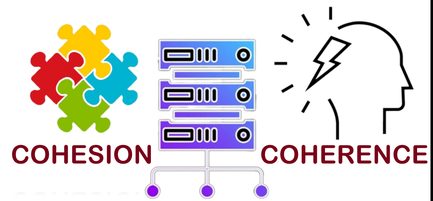
- word2vec semantic similarity
- polysemy words
- hypernymy words

sentiment :

- positive word use
- negative word use
- word polarity

What can we learn from this?

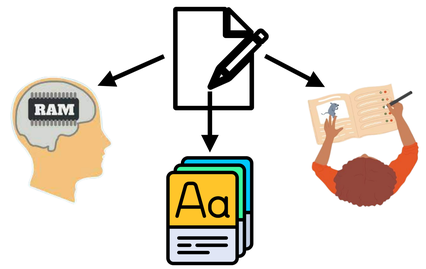
adding to theory:



cohesion used as proxy for the coherence-building processes during comprehension[3]

making predictions:

predict writing quality or model individual differences[4-5]



What does the future hold for NLP?

addressing limitations:

- need for more theory-driven NLP research
- introduce more dynamic/temporal measures
- explore more contexts (e.g., collaborative writing, creative writing, naturalistic dialogue)

educational applications:

- provide adaptive instruction and personalized feedback
- model individual differences
- provide real-time strategy-use suggestions to students

Acknowledgements

This study was supported by grants awarded by the Institute of Education Sciences, US Department of Education (R305A190063 and R305A150193).

References

1. Allen, L. K., Creer, S. D., & Öncel, P. (2022). Natural language processing: Towards a multi-dimensional view of the learning process. In C. Lang, G. Siemens, A. F. Wise, D. Gašević, & A. Merceron (Eds.), The Handbook of Learning Analytics (Vol. 2, pp. 46–53). SoLAR.
2. McNamara, D. S., Allen, L. K., McCarthy, K. S., & Balyan, R. (2018). NLP: Getting Computers to Understand Discourse. In Deep Comprehension. Routledge.
3. Allen, L. K., Mills, C., Perret, C., & McNamara, D. S. (2019). Are you talking to Me? Multi-dimensional language analysis of explanations during reading. Proceedings of the 9th International Conference on Learning Analytics & Knowledge (LAK'19), 116–120. https://doi.org/10.1145/3303772.3303835
4. McNamara, D. S., Allen, L. K., & Crossley, S. A. (2019). WAT: Writing Assessment Tool. In S. Hsiao, J. Cunningham, K. S. McCarthy, G. Lynch, N. Hoover, C. Brooks, R. Ferguson, & U. Hoppe (Eds.), Companion Proceedings of the 9th International Conference on Learning Analytics & Knowledge (LAK'19) (p. 250). SoLAR.
5. Öncel, P., Flynn, L. E., Sonia, A. N., Barker, K. E., Lindsay, G. C., McClure, C. M., McNamara, D. S., & Allen, L. K. (2021). Automatic student writing evaluation: Investigating the impact of individual differences on source-based writing. Proceedings of the 11th International Conference on Learning Analytics & Knowledge (LAK'21), 620–625. https://doi.org/10.1145/3448139.3448207