The Candide Model: A Model for Human-like, Narrative-based Language Understanding

Lara Verheyen

Artificial Intelligence Laboratory, Vrije Universiteit Brussel, Belgium

Abstract. We present the Candide model as a computational architecture for modelling human-like, narrative-based language understanding. The model starts from the idea that narratives emerge through the process of interpreting novel linguistic observations, such as utterances, paragraphs and texts, with respect to previously acquired knowledge and beliefs. Narratives are personal, as they are rooted in past experiences, and constitute perspectives on the world that might motivate different interpretations of the same observations. Concretely, the Candide model operationalises this idea by dynamically modelling the belief systems and background knowledge of individual agents, updating these as new linguistic observations come in, and exposing them to a logic reasoning engine that reveals the possible sources of divergent interpretations. We introduce the foundational ideas underlying this approach and discuss our vision for its large-scale operationalisation.

1 Narrative-based Language Understanding

Today's natural language processing (NLP) systems excel at exploiting the statistical properties of huge amounts of textual data to tackle a wide variety of NLP subtasks. As a result of recent advances in neural machine learning techniques and infrastructure (see e.g. [18, 3]), combined with the availability of huge text corpora, impressive results are now being achieved on many tasks, including machine translation, speech recognition, text summarisation, semantic role labelling and sentiment analysis (for an overview, see [8]). Yet, current NLP systems are everything but capable of modelling human-like, narrative-based language understanding. One reason is that this capacity is hard to cast in the predominant machine learning paradigm. Indeed, human-like narrative understanding is hard to define in the form of an annotation scheme. Narratives are not captured in texts as such, but are construed through an interpretation process. This process is personal, and different individuals may construe different narratives given the same linguistic observations [14]. This diversity in perspectives reflects the richness of human language and cognition, and modelling divergent interpretations constitutes a crucial challenge to the broader computational linguistics community today.

The primary objective of this work is to introduce a novel approach to narrative-based language understanding that starts from the idea that narratives emerge through the process of interpreting novel observations with respect to previously acquired knowledge and beliefs. Concretely, we present a computational model of this interpretation process. The model integrates three main components: (i) a personal dynamic memory that holds a frame-based representation of the knowledge and beliefs of an individual agent, (ii) a construction grammar that maps between linguistic observations and a frame-based representation of their meaning, and (iii) a reasoning engine that performs logic inference over the information stored in the personal dynamic memory. A proof-of-concept implementation of this model is introduced in [16].

Crucially, the representations that result from the language comprehension step take the same form as those stored in the personal dynamic memory. Not only does this mean that these representations can dynamically be merged into the personal dynamic memory to update the knowledge and beliefs of an agent, it also facilitates the use of information stored in the personal dynamic memory to inform the language comprehension process. The information stored in the personal dynamic memory can be queried through a logic reasoning engine, with each answer being supported by a human-interpretable chain of reasoning operations. This chain of reasoning operations explains how the background knowledge and beliefs of an agent guide its conclusions, thereby revealing the narrative construed through the agent's interpretation process.

Personal, dynamic and interpretable models of narrative-based language understanding are of great interest to the fields of computational linguistics and artificial intelligence alike. To the field of computational linguistics, they contribute a perspective that emphasises the individual and contextualised nature of linguistic communication, which contrasts with the static and perspective-agnostic models that dominate the field of NLP today. In the field of artificial intelligence, they respond to the growing interest in the development of artificial agents that combine human-like language understanding with interpretable, value-aware and ethics-guided reasoning (see e.g. [13, 9, 1]).

2 The Candide Model

The model for narrative-based language understanding that we introduce is named after Voltaire's "*Candide ou l'optimisme*" [19]. It is inspired by one of the main themes of the novel, namely that a character's belief system and history of past experiences shape the way in which they interpret the world in which they live.

Following the main theme of the novel, our aim is not to model a single 'true' interpretation of an observation, but to show that different beliefs can lead to different interpretations. Moreover, we consider the belief system of an agent to be dynamic, with the interpretations and conclusions of an agent shifting as more experience and knowledge are gathered. In order to formalise these high-level ideas, we introduce the following operational definitions as well as their technical operationalisations:

Personal dynamic memory (PDM) The personal dynamic memory of an agent is a data structure that stores the knowledge and be-

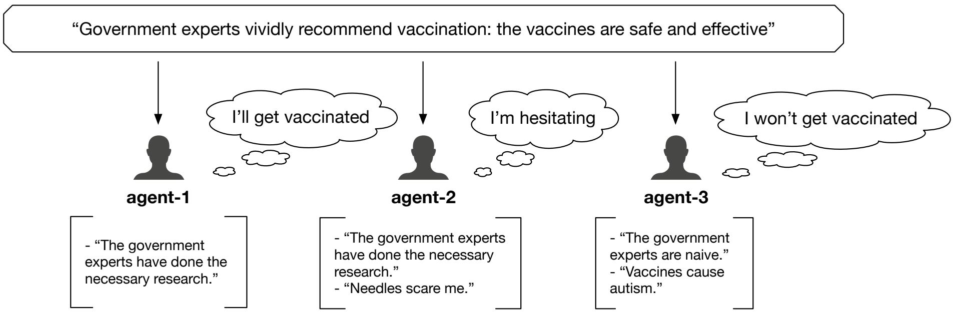

Figure 1. Informal sketch of the Candide model. The model conceives of narrative-based language understanding as the interpretation of a linguistic observation with respect to an agent's individual belief system. Narratives are defined as argumentation structures on the basis of which conclusions are drawn.

liefs of the agent in a logic representation that supports automated reasoning. The PDM is conceived of as a dynamic entity to which new knowledge and beliefs can be added at any point in time. Reasoning over the PDM is non-monotonic, as updated beliefs can alter conclusions. Technically, the PDM consists of a collection of Prolog facts and rules.

Belief system The belief system of an agent at a given point in time equals all information that is stored in the agent's PDM at that moment in time. Each entry in the PDM carries a confidence score, which reflects the degree of certainty of the agent with respect to that entry. However, there exists no formal or conceptual distinction between entries based on their epistemological status, avoiding the need to distinguish between 'knowledge', 'facts', 'opinions' and 'beliefs' for example. We opt for a representation of the beliefs in the form of semantic frames [5, 6]. Semantic frames represent situations, which are evoked by linguistic expressions, along with their participants. For example, the utterance "*Sam sent Robin a postcard*" can evoke a SENDING frame, with "*Sam*", "*Robin*" and "*a postcard*" taking up the roles of SENDER, RECIPIENT and THEME respectively.

Conclusion A conclusion is a piece of information that logically follows from a reasoning operation over the belief system of an agent. A typical example would be the answer to a question.

Narrative A narrative is defined as a chain of reasoning operations that justifies a conclusion based on the belief system of an agent as it is stored in its PDM. Logically, it corresponds to a proof for the conclusion. It is possible that multiple narratives that support the same or different conclusions can be construed by an individual agent. An agent can use the certainty scores carried by the beliefs that constitute its PDM to rank its conclusions and the narratives that support them.

Language comprehension Language comprehension is the process of mapping a linguistic observation to a logic representation of its meaning using the same frame-based semantic representation as in the belief system. The linguistic knowledge needed to support language comprehension is personal and dynamic, and thereby unavoidably constitutes a first layer of individual interpretation. Technically, the language comprehension engine is operationalised using the Fluid Construction Grammar framework [12, 17, 2], which includes a formalism for representing construction grammars, a processing engine that supports construction-based language comprehension and production, and a library of operators for learning construction grammars in a usage-based fashion.

Reasoning The reasoning process is the logic inference that is performed over the beliefs stored in the PDM of an agent when confronted with the meaning of a linguistic observation provided by the language comprehension. As the beliefs stored in the personal dynamic memory of an agent and the meaning of natural language utterances as comprehended by an agent are both represented as a collection of Prolog facts and rules, logical reasoning can naturally be operationalised through SLD-resolution-based inference.

Interpretation The interpretation process comprises all aspects involved in narrative-based language understanding, from the linguistic input to the construction of a narrative that justifies a conclusion. This involves both the language comprehension process, which maps from linguistic input to a logic representation of its meaning, and the reasoning processes that corroborate this meaning representation with the information stored in the PDM, thereby construing narratives that support conclusions.

An informal example of the main ideas underlying the Candide model is shown in Figure 1. Here, three agents observe the same message "*Government experts vividly recommend vaccination: the vaccines are safe and effective*" and are asked whether they plan to get vaccinated. In order to answer the question, the three agents individually interpret this message with respect to the beliefs stored in their PDM and construe a narrative that justifies a conclusion in the form of an answer to the question. The first agent comes to the conclusion that they will get vaccinated, justifying their choice through the narrative that the government experts are competent. The second agent is hesitant to get vaccinated, construing the narrative that vaccines are beneficial but that they are scared of needles. The third agent will not get vaccinated, as they construe the narrative that vaccines are dangerous and that the government experts are being misled.

The example illustrates three properties of narratives that, in our view, constitute crucial challenges in operationalising narrative-based language understanding. First of all, a model of analysis can only be adequate if it captures the personal nature of narratives. Whether or not a conclusion is justified does not depend on its truth or falsehood from an external perspective, but only on whether it is supported by the beliefs held by an agent. Second, narratives are not captured as such in linguistic artefacts. While authors convey messages that are grounded in their belief systems, these messages do not encode the belief systems themselves. Indeed, the intended meaning underlying a message needs to be reconstructed inferentially based on the belief system of the receiver [7]. Finally, it is essential that the interpretation process that is modelled is transparent and human-interpretable. The goal is not merely to draw conclusions given linguistic input, but to reveal the background knowledge, beliefs and reasoning processes that underlie the conclusions that are drawn.

3 Future Outlook

We have defined narratives to be chains of reasoning operations that underlie the conclusions drawn by an individual based on their belief system. This belief system is personal and dynamic in nature, as it is continuously being shaped by new linguistic and non-linguistic experiences. Narratives are thus not captured in texts as such, but need to be construed through a personal interpretation process. A narrative thereby reflects the perspective of an individual on the world.

The construction of a narrative is not a task in itself, but serves the purpose of solving an external task through human-interpretable reasoning processes. As narratives highly depend on external tasks and individual belief systems, they are hard to annotate in linguistic resources. Indeed, whether a narrative is justified or not only depends on whether it is consistent with the input that is observed in combination with the beliefs held by an individual. Narrative-based language understanding therefore largely coincides with the use of explainable methods for solving a variety of NLP tasks, including question answering, text summarisation and sentiment analysis, with the difference that the focus in evaluation shifts from the task accuracy to the soundness of the reasoning processes involved.

The Candide model operationalises this vision through a combination of frame-based constructional language processing and logic reasoning. As such, the belief system of an agent is represented as a collection of facts and rules. The FCG-based language comprehension component is used to map between natural language utterances and a frame-based representation of their meaning. This semantic representation can be integrated as new beliefs into the agent's PDM. The reasoning component can be leveraged to solve tasks by proving logic formulae based on the facts and rules stored in the PDM. It is during this process of logic inference that narratives emerge.

While we have laid the conceptual foundations of a novel approach to narrative-based language understanding, we have left the issue of operationalising the approach on a larger scale unaddressed. We envision an agent to start out as a blank slate, with an empty belief system and grammar. Through experience, an agent would then gradually build up linguistic and non-linguistic beliefs in a constructivist manner through the processes of intention reading and pattern finding. These processes have abundantly been attested in children, see e.g. [11, 15] and have more recently been operationalised at scale in artificial agents through abductive reasoning processes, see e.g. [10, 4, 2]. We consider these preliminary results to be modest yet promising steps towards the moonshot of building personal, dynamic and human-interpretable models of narrative-based language understanding.

4 Conclusion

We introduced the Candide model as a computational architecture for modelling human-like, narrative-based language understanding, thereby presenting an approach that breaks with today's mainstream natural language processing paradigm. Rather than modelling the co-occurrence of characters and words in huge amounts of textual data, our approach focusses on the logic reasoning processes that may justify different interpretations of the same linguistic observations. While this forces us to take an enormous leap back, it bears the promise of contributing a perspective that emphasises the individual and contextualised nature of linguistic communication to the fields of computational linguistics and artificial intelligence.

References

[1] Giulio Antonio Abbo and Tony Belpaeme, 'Users' perspectives on value awareness in social robots', in *HRI2023, the 18th ACM/IEEE International Conference on Human-Robot Interaction*, eds., Maya Cakmak and Iolanda Leite, pp. 1–5, New York, NY, USA, (2023). Association for Computing Machinery.

[2] Katrien Beuls and Paul Van Eecke, 'Fluid Construction Grammar: State of the art and future outlook', in *Proceedings of the First International Workshop on Construction Grammars and NLP (CxGs+NLP, GURT/SyntaxFest 2023)*, eds., Claire Bonial and Harish Tayyar Madabushi, pp. 41–50, (2023).

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, 'BERT', in *Proceedings of NAACL*, pp. 4171–4186. Association for Computational Linguistics, (2018).

[4] Jonas Doumen, Katrien Beuls, and Paul Van Eecke, 'Modelling language acquisition through syntactico-semantic pattern finding', in *Findings of the Association for Computational Linguistics*, eds., Andreas Vlachos and Isabelle Augenstein, pp. 1317–1327. Association for Computational Linguistics, (2023).

[5] Charles J. Fillmore, 'Frame semantics and the nature of language', in *Annals of the New York Academy of Sciences: Conference on the origin and development of language and speech*, volume 280, pp. 20–32. New York, NY, USA, (1976).

[6] Charles J. Fillmore and Collin F. Baker, 'Frame semantics for text understanding', in *Proceedings of WordNet and Other Lexical Resources Workshop, NAACL*, volume 6, (2001).

[7] Paul Grice, 'Logic and conversation', in *Studies in the Way of Words*, ed., Paul Grice, 41–58, Harvard University Press, Cambridge, MA, USA, (1967).

[8] Ivano Lauriola, Alberto Lavelli, and Fabio Aiolli, 'An introduction to deep learning in natural language processing', *Neurocomputing*, **470**(C), 443–456, (2022).

[9] Nieves Montes and Carles Sierra, 'Synthesis and properties of optimally value-aligned normative systems', *Journal of Artificial Intelligence Research*, **74**, 1739–1774, (2022).

[10] Jens Nevens, Jonas Doumen, Paul Van Eecke, and Katrien Beuls, 'Language acquisition through intention reading and pattern finding', in *Proceedings of the 29th International Conference on Computational Linguistics*, eds., Nicoletta Calzolari and Chu-Ren Huang, pp. 15–25. International Committee on Computational Linguistics, (2022).

[11] Julian M. Pine and Elena V. M. Lieven, 'Slot and frame patterns and the development of the determiner category', *Applied Psycholinguistics*, **18**(2), 123–138, (1997).

[12] Luc Steels, 'Constructivist development of grounded construction grammar', in *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04)*, pp. 9–16, (2004).

[13] Luc Steels, 'Personal dynamic memories are necessary to deal with meaning and understanding in human-centric AI', in *Proceedings of the First International Workshop on New Foundations for Human-Centered Artificial Intelligence (NeHuAI@ECAI)*, pp. 11–16, (2020).

[14] Luc Steels, 'Towards meaningful human-centric AI', in *Foundations for meaning and understanding in human-centric AI*, ed., Luc Steels, 5–28, Venice International University, Venice, Italy, (2022).

[15] Michael Tomasello, *Constructing a Language: A Usage-Based Theory of Language Acquisition*, Harvard University Press, Harvard, MA, USA, 2003.

[16] Paul Van Eecke, Lara Verheyen, Tom Willaert, and Katrien Beuls, 'The Candide model: How narratives emerge where observations meet beliefs', in *Proceedings of the 5th Workshop on Narrative Understanding (WNU)*, eds., Nader Akoury, Elizabeth Clark, Mohit Iyyer, Snigdha Chaturvedi, Faeze Brahman, and Khyathi Chandu, pp. 48–57. Association for Computational Linguistics, (2023).

[17] Remi van Trijp, Katrien Beuls, and Paul Van Eecke, 'The FCG Editor: An innovative environment for engineering computational construction grammars', *PLOS ONE*, **17**(6), e0269708, (2022).

[18] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin, 'Attention is all you need', in *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, eds., Isabelle Guyon, Ulrike Von Luxburg, Samy Bengio, Hanna Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett, pp. 6000–6010. Curran Associates Inc., (2017).

[19] Voltaire, *Candide ou l'Optimisme*, Gabriel Cramer, Genève, Switzerland, 1759.