

Narrative Composition: Achieving the Perceived Linearity of Narrative

Pablo Gervás

Universidad Complutense de Madrid, Ciudad Universitaria, 28040 Madrid, Spain

pgervas@sip.ucm.es

The last few years have seen an increased interest in narrative within the field of Natural Language Generation (Reiter et al., 2008; Elson and McKeown, 2010; Siddharthan et al., 2012; Lester, 2012). Narrative is generally acknowledged as a fundamental mode of presenting and communicating information between humans, with different manifestations across media but with a very significant presence in textual form. Yet efforts in Natural Language Generation research have generally side stepped the issue. Aside from the pioneering work of (Callaway, 2002) and an early attempt to bridge the gap between narratology and natural language generation (Lönneker, 2005), the field had mostly avoided narrative until recent times. Two possible arguments may be considered as an explanation of this: one based on the need to restrict initial work within a field to the simpler challenges before tackling the difficult ones, and another based on an assumption that the peculiarities of narrative have already been covered by existing work. Both arguments can be shown to be inappropriate.

With respect to the first argument, the field of natural language generation has for many years operated under the tacit assumption that state of the art technology can only aspire to generating texts within a limited range of domains and genres. These have over the years been defined in different ways, but in spite of changes, literary texts have usually been considered to be outside the range of possible candidates. From an engineering point of view, this kind of restriction made sense when the field was starting, for two important reasons. One, the technological solutions available at the time for the various tasks involved in natural language generation were in their infancy, and the linguistic complexity of literary text might have been beyond their scope. Two, natural language generation arose from a desire to extend the studies that had been carried out for computational analysis of

language to the task of generation, and what was known about language from a computational point of view concerned simple texts. Most of the studies on language and computation had applied similar simplifying assumptions. However, such restricting assumptions are no longer necessary and may be inhibiting progress. In terms of technology, the field has matured significantly over the intervening years. The current state of the art provides a wide range of solutions that may be well suited to address some of the more complex phenomena involved in literary text. Additional objections may be made on the grounds that we do not know enough about these phenomena. Such objections, however valid they might have been originally, are no longer valid either. Many of the phenomena that were considered beyond computational treatment (metaphor, emotion, temporal reasoning, dialogue...) have been the subject of serious and sustained study over the same time period. Many approaches to their computational modelling and treatment have become available. More to the point, the last few years have seen a rise of interest on literary text within the natural language processing community. This is evidenced by the number of workshops addressing topics related to literature: Workshop on Computational Approaches to Linguistic Creativity at NAACL HLT 2009 and 2010, Computational Linguistics for Literature Workshop at NAACL HLT 2012 and 2013, Computational Models of Narrative events held as AAAI Fall symposium in 2010, as LREC workshop in 2012, and as satellite workshop of CogSci 2013, just to name a few.

With respect to the second argument, the recent reappearance of narrative as a research topic for NLG should be enough to dispel the notion that all its problems have already been solved. Narrative has many peculiarities that set it apart from other kinds of text, and the body of work addressing narrative as a research topic within NLG has

at most uncovered and staked out a set of problems and challenges that area waiting further exploration. Of these various open problems in the treatment of narrative, my talk will focus on the problem of narrative composition.

Research on narrative is plagued by the difficulty of establishing a definition of the term that is both sufficiently formal to act as foundation for scientific rigour, and sufficiently rich to cover the fundamental aspects that people associate with the term. At the present stage of development, tentative definition need to be established, to be later confirmed on the basis of empirical work and successful evaluation of results. The talk will outline some of the possibilities that must be considered (arising from established definitions in the field of narratology) and some of the restrictions that arise from the computational nature of the task. From the combination of these constraints, a working model of narrative structure will be outlined. However, it is clear that such a model must document the relation between a semantic description of the content of the narrative (what is usually termed the *fabula*) and its rendition as a sequential discourse. The task of *narrative composition* will be specified as the task of constructing such a discourse (or discourse plan) for a given semantic description of fabula. This discourse should be susceptible of being converted into text and it should appropriately convey the set of events in the fabula in such a way that satisfies a number of traditionally accepted requirements (like having an identifiable theme, a certain temporal and causal coherence, a recognisable set of characters...). A number of basic narratological concepts will be described where they provide tools for breaking down the task into computationally tractable subproblems. Of particular interest is the concept of *focalization*, which is used by narratologists to describe the way certain segments of a narrative follow a particular character, and which provides a useful computational representation of both the granularity and the shift in focus employed during the process of converting the semantics of the fabula into a linear discourse.

As part of the talk, narrative composition will be framed in terms of the accepted task breakdown for natural language generation, considering that it may involve a combination of content determination and discourse planning that cannot be segregated into separate subtasks. The talk will also

discuss the relation of the task of narrative composition with a number of existing research problems such as story generation (which could correspond to the construction of fabula but is sometimes simplified down to construction of a discourse directly) and creativity (which has been addressed with respect to story generation but may also constitute a fundamental ingredient of the composition task).

Acknowledgments

The work on which this talk is based was partially supported by the Ministerio de Educación y Ciencia (TIN2009-14659-C03-01).

References

- Charles B. Callaway. 2002. Narrative prose generation. *Artificial Intelligence*, 139(2):213–252.
- David K. Elson and Kathleen R. McKeown. 2010. Tense and aspect assignment in narrative discourse. In *Proceedings of the 6th International Natural Language Generation Conference, INLG '10*, pages 47–56, Stroudsburg, PA, USA. Association for Computational Linguistics.
- James Lester. 2012. Expressive nlg for next-generation learning environments: language, affect, and narrative. In *Proceedings of the Seventh International Natural Language Generation Conference, INLG '12*, pages 2–2, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Birte Lönneker. 2005. Narratological Knowledge for Natural Language Generation. In Graham Wilcock, Kristiina Jokinen, Chris Mellish, and Ehud Reiter, editors, *Proceedings of the 10th European Workshop on Natural Language Generation (= ENLG 2005)*, pages 91–100, Aberdeen, Scotland, August.
- Ehud Reiter, Albert Gatt, François Portet, and Marian van der Meulen. 2008. The importance of narrative and other lessons from an evaluation of an nlg system that summarises clinical data. In *Proceedings of the Fifth International Natural Language Generation Conference, INLG '08*, pages 147–156, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Advait Siddharthan, Matthew Green, Kees van Deemter, Chris Mellish, and René van der Wal. 2012. Blogging birds: generating narratives about reintroduced species to promote public engagement. In *Proceedings of the Seventh International Natural Language Generation Conference, INLG '12*, pages 120–124, Stroudsburg, PA, USA. Association for Computational Linguistics.