Digital discourse analysis and Wikipedia: Bridging the gap between Foucauldian discourse analysis and digital conversation analysis

Eva Gredel

Chair of German Linguistics, University of Mannheim, Schloss, Ehrenhof West, Raum 253, 68131 Mannheim, Germany

Abstract

In the last decade, analyzing online interaction rapidly gained in popularity and especially Wikipedia with its collaborative nature is a favored research object in linguistic studies. In discourse (linguistic) studies, there are some studies focusing on Wikipedia's encyclopedic article pages and others concentrating on conversational-like data on Wikipedia's talk pages. While the former can be located in the paradigm of Foucauldian discourse analysis, the latter have a rather conversation analytic approach. This paper gives an impulse to consider the convergence of data – namely the text-text convergence of article and talk pages on Wikipedia – by integrating digital conversation analysis (Giles et al., 2015) with Foucauldian discourse analysis (Spitzmüller and Warnke, 2011a). The paper begins to address this gap by conducting a detailed microanalysis on how discussions on talk pages impact on regimes of meaning on Wikipedia's article pages. The discourse fragment under investigation is the controversially discussed entry *Annexation of Crimea by the Russian Federation*. The considerations result in an analytic framework that integrates Foucauldian discourse analysis and digital conversation analysis. Therefore, this paper contributes to the discussion about digital conversation analysis by investigating digital discourses on Wikipedia. The microanalysis reveals communicative strategies and practices of interactivity on article pages and on corresponding talk pages. The peculiarities of Wikipedia's data will be explained, modes of analysis discussed and the challenges of the suggested integration of methods will be explored.

1. Introduction

Wikipedia is not only a controversially discussed reference system that is used by many as a first source, but it is also one of the most successful projects in Web 2.0 environments. Currently, Wikipedia contains 38,000,000 entries (Wikipedia, 2016a) and it scores highly in the category "The top 500 sites on the web" of the Alexa ranking (rang 6, Alexa, 2016). The main goal of users on Wikipedia is to contribute to a high-quality online encyclopedia in a peer-production process. Therefore, the MediaWiki software used in the project offers various opportunities to support collaborative content production: revision histories document each change made to a wiki page and corresponding digital discussion spaces, particularly the so-called talk pages, serve to coordinate and discuss project issues. Given this nature of the online encyclopedia, it is a prime example for digital environments enabling online interaction and thus a complex and worthwhile data source for linguistic and cultural studies. Consequently, Wikipedia can be seen and studied in many ways: as a case of collaborative authorship (Emigh and Herring, 2005), as a news source (Bucher and Büffel, 2006, p. 137), as a hypertextual space (Storrer, 2008, 2012), as a cultural reference (Rogers, 2013), and as an (online-) encyclopedia (Haß, 2016).

E-mail address: eva.gredel@phil.uni-mannheim.de.

In this paper, Wikipedia is investigated as a vast multilingual corpus that sheds light on digital discourses. Not only article pages but also talk pages are relevant to this approach; article pages are products of discourses with a rather encyclopedic style whereas talk pages contain conversation-like data revealing the process of negotiation in collaborative content-production. The findings of this paper indicate that both spheres are highly interconnected via links confronting researches with a great variety of different linguistic data. But, existing discourse analytic studies on Wikipedia in the tradition of Foucault (Pentzold, 2007; Beyersdorff, 2011; Arendt and Dreesen, 2015) concentrate on the encyclopedic part of Wikipedia or neglect the conversation-like nature of digital data on talk pages. By focusing on the aforementioned convergence of data on Wikipedia, this paper aims at addressing this research gap to contribute to methods and methodologies enabling research on discourses in digital media.

The Wikipedia article *Annexation of Crimea by the Russian Federation*, the corresponding talk page and revision history as well as policies on Wikipedia constitute the discourse fragment under investigation. Section 2 delivers a detailed and systematical description of the Wikipedia's data as a hypertext and the interdependencies of its namespaces[1] in the light of a discourse analytic inquiry. Section 3.1 puts the argument forward that analyzing Wikipedia in the discourse analytic tradition of Foucault is promising and helps to overcome the "newspaper bias" (Warnke, 2013, p. 191) of Foucauldian discourse analysis. Restrictions of the framework DIMEAN (multi-layer approach to discourses, Spitzmüller and Warnke, 2011a) in the face of complex digital online data are outlined. Consequently, aspects of a digital conversation analysis as discussed by the network Microanalysis of Online Data (Giles et al., 2015) are introduced in section 3.2 and are applied to the aforesaid type of data putting a special focus on the hypertextual features of Wikipedia. In section 3.3 a new analytic framework is presented to meet the needs of digital discourse analysis bridging the gap between Foucauldian discourse analysis and digital conversation analysis. Section 4 delivers a detailed analysis where special data and phenomena of digital discourses on Wikipedia are explained and section 5 contains a discussion and closing remarks.

2. Data: convergence of media and convergence of data

In her paper on discourses in Web 2.0 environments and on CMCMD (convergent media computer-mediated discourse) Susan Herring elaborates on the fact that "collaborative text production of the sort that takes place on Wikipedia represents a new kind of online discourse" (Herring, 2013, p. 15). The one particular feature of such discourses is the convergence of media and consequently the convergence of various modes such as pictures and text or video and text. In the case of Wikipedia Herring argues: "Less prototypically (because it involves the convergence of text with text rather than the convergence of text with another mode), CMCMC [convergent media computer-mediated communication] is also illustrated by [. . .] 'talk' pages associated with Wikipedia articles" (Herring, 2013, p. 5).

The mentioned text-text convergence on Wikipedia is enabled by links that connect not only the article pages (Fig. 1A) with corresponding talk pages (Fig. 1B), as Herring proposes, but also an interface for editing the article page (Fig. 1C), the revision history (Fig. 1D), as well as various other namespaces (e.g. user talk pages or policy pages). The variety and peculiarities of these converging data provoke many issues for computer-mediated discourse analysis: some studies have featured either article pages or talk pages, but the convergence of data has received little attention to date. From a methodological perspective, the main challenge is that there are not only "new types of content" (Herring, 2013, p. 5) but also new types of born-digital data to be analyzed such as links, revision histories, and timestamps.[2]

For instance, the recent incidents described by the article *Annexation of Crimea by the Russian Federation* (Fig. 1E) is distinguished from the historical event of 1783 that can be found under the lemma *Annexation of Crimea by the Russian Empire* and that is available via another link (Fig. 1G). Fig. 1 also documents the discourse analytically interesting fact that Wikipedia reduces ambiguity of terms and entities by allowing users to establish redirects[3]; when searching the lemma *Crimean Crisis* – a term that was also coined in the media to denote incidents in Ukraine in 2014 – the user is automatically redirected to the lemma *Annexation of Crimea by the Russian Federation* (Fig. 1F). The example at hand further demonstrates that Wikipedia is not only characterized by text-text convergence but also by text-picture-convergence (Fig. 1H) and more and more text-picture-video convergence making it an even more challenging multimodal environment (Storrer, 2008).

The convergence of data is further illustrated by the previously mentioned case of the incidents in Crimea in 2014 that constitute a controversially discussed topic in media discourses with international relevance. When visited on April 2016, the user found the article page *Annexation of Crimea by the Russian Federation* with a page length of 220,907 bytes

---

[1] A namespace is "a virtual container for different types of content on the wiki; namespaces are defined by different prefixes, such as *talk:* or *Wikipedia:*, which appear before page names; articles are in the main namespace" (Ayers et al., 2008, p. 473).

[2] A timestamp on Wikipedia is the sequence of information that records the time at which an edit is done.

[3] A redirect is a page created so that navigation to a given title takes the reader directly to a different page.