

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/67115> holds various files of this Leiden University dissertation.

Author: Lukac, M.

Title: Grassroots prescriptivism

Issue Date: 2018-11-22

From usage guides to Wikipedia: Re-contextualising the discourse on language use¹

5.1 Introduction: Expert discourses on language use

The word *usage* in linguistic contexts has two different attitudinal connotations, one neutral and the other more judgemental and negative (Peters, 2006, pp. 759–60; Busse & Schröder, 2009, p. 72). In general terms, *usage* refers to the customary or habitual way of doing something, but in the context of linguistic prescriptivism, *usage* may refer to linguistic practices that are contrasted with what is prescribed, namely the rules of the standard language (Allen, 1992, p. 1071). Such non-standard usage is often labelled ‘bad usage’ (Allen, 1992, p. 1071), ‘bad grammar’ or simply ‘a mistake’ (Bloomfield, 1944, p. 45), and it may be stigmatised in the prescriptive tradition. Linguistics as a discipline primarily aims at describing the rules of use accurately or, in the words of Charles Fries, linguists traditionally hold that ‘there can be no *correctness* apart from usage’ (as cited in McArthur, 1992, p. 421).

¹ Lukač, M. (2017). From Usage Guides to Wikipedia. Re-contextualising the discourse on language use. In M. Bondi, S. Cacchiani and D. Mazzi (eds.), *Discourse In and Through the Media: Recontextualising and Reconceptualising Expert Discourse* (pp. 315–50). Newcastle upon Tyne: Cambridge Scholars Publishing.

Quite contrary to these views originating from structural linguistics, the prescriptive tradition introduces evaluative judgements by prescribing certain rules for usage and proscribing others (McArthur, 1992, p. 446). The history of linguistic prescriptivism in the English language goes back to the eighteenth century (Beal, 2009), when the prescriptive rules were established as a by-product of the early grammarians' attempts to codify English grammar (Peters, 2006, p. 761). In this period prescriptive rules such as the rules proscribing against double negation and split infinitives were first introduced. The history of prescriptivism in the English language is recorded and preserved in the tradition of the usage guide genre or the 'para-lexicographic tradition' (Peters & Young, 1997, p. 317), which continues to persist independently in spite of the 'descriptive turn' in lexicography and in the writing of grammar books. The 'descriptive turn' has been greatly influenced by the establishment of linguistics as a discipline and, subsequently, by using naturally occurring data for studying language. The language advice in usage guides is still, however, for the great part the result of the topical selection, acceptability judgements and attitudes of their authors (Algeo, 1991, p. 6; Peters & Young, 1997, p. 317); in other words, usage guides are often subjective and dependent on introspection.

Usage guides, however, are not the only records of the prescriptive tradition. The history of the standard language ideology, the consciousness of the standard and of 'correct' and 'incorrect' language use (Milroy & Milroy, 2002, p. 25) is charted out in the 'complaint tradition' (Milroy & Milroy, 2002, Chapter 2), which consists of public

complaints about the misuse of language and about linguistic decline, which are commonly published in letters-to-the-editor sections of newspapers and, more recently, on weblogs and Internet forums. Usage advice has also found its place in the new media genres, on specialist weblogs, wikis, and in various types of social media (Pinterest boards, Facebook groups, Twitter accounts, etc.) dedicated to usage (Schaffer, 2010).

Although both the usage guide tradition and the complaint tradition serve to maintain the standard language ideology, they traditionally represent different groups of participants in the discussions on linguistic prescriptivism. On the one hand, there are the usage guide authors, the prescriptivists, and on the other the members of the general public, popularly known as language pedants or ‘language mavens’ (Cameron, 1995, p. vi). Since the introduction of Web 2.0, however, the two groups, the ones engaging in giving usage guide advice and the members of the general public, are no longer clearly separated. In the medium where publishing became accessible to anyone with an Internet connection, many of the members of the general public with an interest in usage got the opportunity to create their own weblogs and contribute to discussions on language use (Schaffer, 2010, pp. 23–4). One such medium where language use is discussed and described is Wikipedia, the online collaborative encyclopaedia community. In the sphere of advice on language use where linguistic authorities traditionally functioned as gatekeepers, Wikipedia currently functions as a platform for translation between the groups that would otherwise not communicate. The status of grammars and usage guides as authorities is thus chal-

lenged and the questions of language use are negotiated on a more equal footing between the language experts and the general public.

In this paper I present an analysis of Wikipedia entries and the secondary Wikipedia Talk pages, which editors use to discuss the respective Wikipedia entries, on usage items such as *ain't*, the split infinitive, preposition stranding, and *who/whom*. The reason behind choosing this particular online genre for the analysis of the discussions on and descriptions of language use is the fact that all of the content is created by the members of the general public who negotiate the content of the entries in the secondary Talk pages. On the other hand, the reason for choosing language use as a topic for demonstrating the phenomenon of re-contextualisation of expert discourse is the afore-mentioned established gap between the advice-giving experts and the advice-seeking laypeople.

Although Wikipedia instructs its contributors, 'Wikipedians', to provide informative and descriptive accounts of usage items, the Talk pages demonstrate that many of the contributors express prescriptive attitudes towards usage. I focus here on an analysis of the prescriptive and descriptive arguments of the collaborators creating Wikipedia entries and the guiding principles of Wikipedia that provide the basis for the construction of the entries.

The second point of the analysis focuses on a comparison of the Wikipedia entries on selected usage items with their equivalent entries in usage guides, which are retrieved from the Hyper Usage Guide of English or HUGE database (developed by Robin Straaijer at Leiden University). The HUGE database (Straaijer, 2014) is a growing collec-

tion of usage guides covering the period from 1770 onwards. The database currently includes 77 usage guides and its aim is to combine a history of usage advice into a single library. The HUGE database has been envisioned within a research project at Leiden University Centre for Linguistics called ‘Bridging the Unbridgeable: Linguists, Prescriptivists and the General Public’. In doing so, I will address in section 5.5 the differences in the styles of the Wikipedia and usage guide entries, analyse the arguments provided, and the language of prescription and description.

5.2 The usage guide as a genre

The usage guide has been described as ‘a neglected genre’ (Weiner, 1988, p. 171), although a rising number of studies have been devoted to it since the late 1980s (Busse & Tieken-Boon van Ostade, 2011). The usage guide genre has been defined as ‘an integrative all-in-one reference work written for educated lay people that bridges the traditional divide between a grammar and a dictionary’ (Busse & Schröder, 2009, p. 72). Being a usage guide author himself, Weiner (1988, p. 173) defines the goal of usage guides as helping its users decide between alternatives which from a descriptive point both exist in language, but of which for some reason or another one is considered less good English than its alternative.

Usage guides are not intended for the language learner, but rather for the native speaker (Weiner, 1988, p. 173). Among the native speakers, the target readership of such usage guides are in Labovian terms ‘the linguistically insecure’ (Beal, 2009, p. 42), social climbers,

who are not ‘born into’ using the standard language, and who are ‘shamed by their English’ (Beal, 2009, p. 42). As the author of the arguably most influential usage guide, *Modern English Usage*, Henry Fowler argues in his correspondence with the Oxford University Press:

In point of fact we have our eyes not on the foreigner, but on the half-educated Englishman of literary proclivities who wants to know Can I say so-&-so?, What does this familiar phrase or word mean?, Is this use English? (...) the kind of Englishman who has idioms floating in his head in a jumbled state, & knows it... (as cited in Burchfield, 1991, p. 96)

Usage guides continue to be extremely popular; a study by Busse and Schröder (2008) showed that the numbers of the usage guide publications are steadily on the rise, presumably along with the rising popularity of other guides, self-help and how-to literature.

Several linguists have provided critical accounts of the usage guide tradition. In his classification of different types of usage guides, Algeo (1991, pp. 6–13) points to the fact that the largest group of usage guides is that consisting of books that largely depend on *ipse dixit* judgements, in other words, guides that largely rely on the personal judgements of their authors (Algeo, 1991, p. 6). Usage guides on the other side of the prescriptivism-descriptivism spectrum are far and few between. An example of a more descriptive usage guide according to Algeo is *Webster’s Dictionary of English Usage*, which he describes as ‘a book [which] does not tell people what they ought to say, but explains the options and the likely consequences of choosing one option over another’ (Algeo, 1991, p. 11). Although there seems to be a chronological shift towards usage advice that is more explanatory and usage-

based, some more recent publications, such as Burchfield (1996) and Garner (1998), still seem to be relying on personal, subjective judgements of their authors (Peters, 2006, p. 765). Although usage guides tend to vary considerably in the choice of their items, they also usually include the traditional shibboleths of usage (Peters, 2006, p. 12) that are also known as ‘old chestnuts’ (Weiner, 1988, p. 173). It can be argued that normative judgements essentially belong to usage guides. Their readers expect clear, user-friendly guidance, which is usually lacking in the more objective accounts of the more jargon-loaded grammar books and dictionaries (Busse & Schröder, 2009, p. 84).

Despite the fact that usage guides are read for their often clear-cut advice, scholars analysing the tradition of usage guides warn of their lack of lateral referencing (Peters & Young, 1997, p. 318), which serves little use to the contemporary reader who would like to be informed about current usage trends. Those works that do not include the analyses of contemporary usage do little more than replicate conservative attitudes, support ‘the paralexigraphic tradition’ and institutionalise the tradition of ‘tertiary responses to language’, which are widely accepted regardless of their validity (Bloomfield, 1944, p. 45).

5.3 The history of collaboration in knowledge creation: From the *OED* to Wikis

Collaboration in knowledge creation, as we find it today in Wikipedia, is hardly a novelty: it has been around since biblical times when scribes simultaneously edited, updated, interpreted, and reinterpreted texts as they were transcribing them (McArthur as cited in Stvilia et al., 2008, p.

983). One of the greatest global processes of collaboration and co-creation in the pre-computer age, which is also of special interest for linguistics, is that of the making of the *Oxford English Dictionary* (*OED*) (Simpson, 2004, pp. 192–196; Stvilia et al., 2008, p. 983; Bhal-la, 2011, pp. 8–9).

Besides the practice of contacting specialist consultants, the *OED* has a long history of recruiting volunteer contributors from the members of the public (Simpson, 2004, pp. 193–4).² In 1879, the then recently appointed editor of the *OED*, James Murray, decided to instigate *Appeal to the English-Speaking and English-Reading Public to Read Books and Make Extracts for the Philological Society's New Dictionary* (Mugglestone, 2005, p. 15). This ‘crowdsourcing’ process contributed significantly to the *OED*, with thousands of contributors and millions of archived physical slips which have been in use until today (see the *OED*'s website). Not unlike Wikipedia, Murray's *Appeal* encouraged democratic collaboration, which allowed everyone to take part: ‘This is work in which anyone can join, even the most indolent novel-reader will find it little trouble to put a pencil-mark against any word or phrase that strikes him, and he can afterwards copy out the context at his leisure’ (as cited in Mugglestone, 2005, p. 16). In order to manage the work done by the volunteers more efficiently, Murray complemented the initial *Appeal* by a pamphlet including a more targeted approach including ‘lists of wants’ and ‘desiderata’ which made the collaborative process more helpful for the editors (*OED*'s website). The collaboration

² The history of the *OED* website is documented on the *OED* website <http://public.oed.com/the-oed-appeals/history-of-the-appeals/>.

of the *OED* with the general public has continued until today in the form of online appeals.³

Although quite ground breaking, the collaborative participation on the *OED* did include organisational difficulties: the contributions were occasionally obsolete, incorrect or duplicated, and they required a substantial amount of editing and assessment from the central institution. Such issues today have been largely overcome with the development of collaborative processes that are greatly facilitated by the introduction of online content management technologies such as Wikis (Stvilia et al., 2008, p. 984), through which subsequently thousands of contributors continue creating dictionaries (*Wiktionary*) and online grammars and usage guides (*English Grammar and Usage Wiki*).

5.4 Wikipedia: The online collaborative encyclopaedia community

As of April 2014 Wikipedia is the sixth most visited website in the world.⁴ It has become the most widely used tool for knowledge dissemination and the largest collaborative text-editing project in the history of human kind. The multilingual and freely accessible online encyclopaedia is available in 285 languages. Potentially every Internet user can edit documents on Wikipedia. To date it has approximately 31 million articles and 76,000 active contributors. Contributors are considered to be 'active' with a minimum of five contributions per month. There are currently 4,518,174 articles written in English.⁵ The greatest contribution of this project is the fact that through it a vast quantity of infor-

³ See 'OED Appeals' www.oed.com/appeals.

⁴ See 'The top 500 sites on the web' www.alexa.com.

⁵ See 'Wikipedia: About' (<http://en.wikipedia.org/wiki/Wikipedia:About>).

mation that was previously accessible only through traditional knowledge institutions such as libraries has been greatly popularised and made available to the general public.

Wikipedia is an instantiation of the Wiki software concept, which allows users to collaborate in a web-based manner and to edit a single document. Originally it was developed by Ward Cunningham, who was looking for a tool that would enable better collaboration among developers (Ebersbach et al., 2008, p. 14). Cunningham also introduced ‘wiki-philosophy’, which is based on the unlimited possibility to create and edit pages, and is referred to as the ‘open editing concept’ (Fichter, 2005, p. 47). Wiki systems are also document management systems which allow users to trace back every single change that has been made to the document and reverse it. The open editing concept was designed to enable a more democratic access to information systems to users with no advanced technological skills.

Since Wikipedia is constructed collaboratively, it can be regarded as an online community, which is defined as a group of people ‘who come together for a particular *purpose*, and who are guided by *policies* (including norms and rules) and supported by *software*’ (Preece & Maloney-Krichmar, 2005). Bruns (2008) places Wikipedia in the wider context of the social media, which aim at participation by a wider audience, not just by a community of experts. Through client-based editing, the traditional linear knowledge dissemination process from the expert to the lay community has been enriched by the fact that laypeople and experts are now interacting on the same platform in a multilateral way. The representation of knowledge on Wikipedia is constructed in a self-

organised way; nevertheless, Wikipedia is not an egalitarian system (Gutounig, 2015, pp. 149–150). Established hierarchies and powers are, however, not attributed through external status, but through actual contribution to the system. In this way they can be defined as ‘meritocracies’ (Bruns, 2008, p. 25).

The entries and the editing interface are just the most visible side of Wikipedia. To enable the discursive aspect of collaboration, Wikis usually have discussion functions, in the case of Wikipedia, in the form of Talk pages. These pages enable not only collective editing of the entries, but also engaging in a discussion on the topic of the entry. The Talk pages should lead to a usually temporary consensus regarding the entry in question. The basic principles of the open-editing concept combined with the traceability and discussion functions enable the self-organised editing process without central governance among people who are usually not acquainted with each other. Due to these characteristics, Wikipedia seems to establish what is considered in some aspects to be an unprecedented, emergent discourse context (Herring, 2013, p. 14).

5.4.1 Related work

Due to the success of the Wikipedia project, a substantial number of scholarly publications have dealt with the phenomenon in the course of the last decade. Studies have so far focussed on the collaboration and coordination patterns in Wikipedia (Viégas et al., 2004). Researchers have devoted attention to the acts of vandalism on Wikipedia, i.e. edits that were made with bad intentions (Viégas et al., 2007, p. 3; Potthast et

al., 2008), as well as to the quality of Wikipedia entries when compared with traditional encyclopaedias (Giles, 2005). These studies showed that entries in *The Encyclopaedia Britannica* and Wikipedia go head to head concerning the number of factual errors or misinterpretation of important concepts.

In linguistics, Emigh and Herring (2005) were among the first to analyse the aspect of text production on Wikipedia. By performing genre analysis on the level of formality and informality of Wikipedia entries, they found that in spite of the collaborative and open-editing approach of Wikipedia, its level of formality can be compared to that of traditional encyclopaedias. Myers (2010) devoted a book to the analysis of the discourse of blogs and Wikis. According to Myers (2010, p. ix), there are two main reasons to analyse Wikipedia linguistically; the first is focussing on new aspects of the language of emerging web genres, and the second is linguists' contribution to a phenomenon that has been dealt with primarily in the field of technology. In his analysis of Wikipedia, Myers primarily focussed on the interaction among 'Wikipedians' on Talk pages.

5.4.2 The structure of Wikipedia entries and Talk pages

The fact that Wikipedia entries hardly differ from the entries in traditional print encyclopaedias can be explained by the phenomenon that, since the beginnings of the project, Wikipedians 'were guided by the rhetorical models of existing encyclopaedias' (Shirky, 2010, p. 116) and that they have 'internalised cultural norms of encyclopaedic style' (Herring, 2013, p. 15). The formality and stylistic homogeneity of en-

tries (Herring, 2013, p. 9) is for a large part the accomplishment of the active rank-and-file Wikipedia users, who continue changing existing text according to the stylistic norms of a traditional encyclopaedia, and who are also referred to as page ‘watchers’ (Viégas et al., 2004, p. 580). A relevant point of Wikipedia’s departure from traditional print encyclopaedias are its secondary Talk pages, which greatly determine knowledge construction processes on Wikipedia, and which are an essential part of defining a collaborative online encyclopaedia. An example of such a Talk page can be found in Figure 5.1.

Article Talk

Talk:Preposition stranding

From Wikipedia, the free encyclopedia

Contents (hide)
1 "Controversial" stranded preposition in German?
2 Minor Edit
3 Arguments for and against?
4 Alternatives
5 Expansion and cleanup.
6 Preposition stranding in non-standard French
7 Winston Churchill's quote
8 There is an object
9 some issues...
10 Theoretical Bias?
11 Il faudra agir selon ("We'll have to act accordingly")
12 is this...?
13 Globalize
14 Certain prepositional passives
15 "Out of" a preposition?
16 Prepositional Endings in English Usage Guides
17 Prepositions immediately following their object

Figure 5.1 Talk page on Preposition stranding

Talk pages are devoted to the discussion of issues surrounding the topics on ‘real’ pages. They provide a different forum from the Wikipedia entries, and they consist of discussions on what information should or should not be included in the main articles (Viégas et al., 2004, p. 576). The Talk pages on Wikipedia contribute to Wikipedia being a successful system. They serve a number of functions in creating and managing

articles, such as strategic planning of edits, enforcement of Wikipedia policies and conducting guidelines (Viégas et al., 2007, p. 1). Talk pages have a central role in keeping up the quality of Wikipedia entries. They are conversation places and as such are governed by different rules from the entries. For example, the entries are not deleted in the Talk pages, and this is often the case with the edited content in the entries, and the participants are encouraged to sign their postings (Viégas et al., 2007, p. 6). There are three main layout and mark-up conventions in Talk pages, (1) signatures, (2) indentation and (3) discussion topics (Viégas et al., 2007, pp. 6–7). The convention for the contributors to sign their postings in Talk pages is respected in 67% of the instances (Viégas et al., 2007, p. 7). When the postings are left unsigned, only an IP address appears in the end of the posting. Users usually indent their answers to previous postings in order to visually disambiguate the relationship between the postings. Finally, contributors are instructed to put each new conversation topic at the end of the existing Talk page.

Researchers have found that the genre of Talk pages most resembles informal web discussion boards (Emigh & Herring, 2005, p. 7; Myers, 2010, pp. 154–56), which is in contrast to the formal style of the entries themselves. The linguistic features identified as indicative of the informal, web-chat style are first person pronouns, contractions, emoticons, and informal lexicon (Emigh & Herring, 2005, p. 8) as well as conversational discourse markers, such as discourse particles (*well*, *umm* and *ahem*), nonwords (*ahem*, *uh huh*) and politeness markers (modals, verbs of cognition and perception, and modal adverbs) (Myers, 2010, Chapter 10). Politeness plays a crucial role in Talk pages, as

it softens potential conflicts and contributes to the general feeling of belonging to a Community of Practice (Myers, 2010, pp. 155–56).

Concerning the topics and functions of discussions on Talk pages, Viégas et al. (2007, pp. 7–8) analysed the dimensions along which contributions to Talk pages can be classified. Wikipedians most commonly use Talk pages to request for coordination of the entry edits and they also approach the participants as a community of experts and ask for information. Talk pages occasionally include off-topic remarks, which means that the participants discuss topics loosely related to the entry, or rather report on their own experiences and opinions. Some of the Talk pages provide insights into ‘edit wars’ between groups of Wikipedians, in which two people or groups of opposing opinions alternate between versions of the page. In some instances of discussions on Talk pages, participants also point to internal resources, namely, other Wikipedia pages (Viégas et al., 2007, pp. 8–9). Ferschke et al. (2012) analysed dialogue acts in Simple English Talk pages, and they found that the most common types of entries on Talk pages are information-providing comments, in which Wikipedians communicate new information, request information or suggest changes. Wikipedians tend to report on their edits in order to justify the changes made to the entries, and almost 40% of turns in Talk pages are article criticisms. It is common to start a discussion or a topic on a Talk page by referring to a particular deficiency in the accompanying article.

Myers (2010, pp. 146–154) found in the analysis of types of argumentation on Wikipedia that the rhetoric of Wikipedians is greatly influenced by the Wikipedia’s explicit principles: (1) Neutral Point of

View, (2) No Original Research, (3) Verifiability, (4) Be bold and (5) Civility. Neutral point of view, or NPOV as Wikipedians refer to it in their discussions, implies that editors should try to include a full range of views on a topic in their accounts. No Original Research (NOR) warns editors against publishing new ideas or mentioning facts that cannot be documented instead of referring only to the available sources. Additionally, every statement needs to be verified; in case it is not, editors insert ‘Citation needed’ tags into the entries. The ‘Be bold’ principle of Wikipedia urges editors to rather say something roughly accurate than to say nothing at all, and to build up content. Finally, due to the fact that the goal behind each entry is to reach consensus, editors are urged to be polite and cooperative.

5.5 Analysing Wikipedia entries on language use

For this study I selected seven Wikipedia entries on usage items, in alphabetical order: *Ain't*, *Double negative*, *Fewer vs. less*, *Gender neutrality in English*, *Preposition stranding*, *The Split infinitive* and *Whom* (Table 5.1). These entries were selected as they all included accounts of acceptability of usage and were described in the context of the prescriptivism-descriptivism debate.

Most of the included usage items have a longstanding place in the history of the prescriptive tradition. The prescriptive rules regarding the respective usage items could be summed up as following: Do not use *ain't* when you mean *isn't* or *aren't*. Do not use more than one negative particle to negate the same clause. Do not use *less* instead of *fewer* with plural countable nouns. Do not use gender-specific words in non-gender

specific contexts. Do not place a preposition in the end of the sentence, but before the noun phrase it modifies. Do not insert anything between the infinitive *to* marker and the verb-form itself. Use *whom* as the objective form of the interrogative pronoun *who*.

In Table 5.1, I provide data on the selected Wikipedia entries, including their length, number of watchers who are alerted when changes are made to each of the entries, date of page creation, total number of edits, total number of distinct authors, page views, and comparison of the data with average values for Wikipedia entries where these values are available, namely for page length, number of edits and page views. All pages on usage items are considerably more often edited when compared to an average Wikipedia entry (see column Total N of edits), which is on average edited 21.82 times. The usage entries attract more attention, contribution and updates from the editors. The entries on usage items are also considerably longer than an average entry. Although they are heavily edited, not all entries on usage items are viewed more than the average (see column Difference to avg. page views).

Preposition stranding and *Whom* are the least popular among the selected usage entries. The *Split infinitive* and *Ain't* attract by far the most readers. Another sign of popularity are the numbers of editors of a particular entry, which are the highest for *Gender neutrality*, *Double negative* and the *Split infinitive*. In the following sections (5.5.1) and (5.5.2), I give a description of the Wikipedia contributors and I describe the entries on usage items themselves.

5.5.1 *The editors*

Researchers usually distinguish between two general groups of Wikipedia editors (Bryant et al., 2005), novice and expert users. Novices contribute to topics they are familiar with, identifying omissions and weaknesses, and correcting them. Up to 60% of Wikipedia's registered users never make another edit after their first 24 hours.⁶ The experts' goals expand as they continue contributing to Wikipedia. Although they continue to improve the quality of the content, they have an additional role in the community, as they become concerned with maintaining the quality of Wikipedia itself. Expert editors use tools such as watchlists and WikiProjects, which allow them to review changes and observe instances of vandalism.

Another relevant distinction between novice and expert users is their sense of community. Whereas for novices Wikipedia seems to be a collection of articles, more experienced contributors view themselves as members of a tribe in which they establish their identity through their contributions (Bryant et al., 2005, p. 7). Novice users usually become experts through observation and direct coaching by the more knowledgeable users (Bryant et al., 2005, p. 8).

In this study I additionally attempted to address the question: 'Who are the contributors to the entries on usage items?', and more specifically, 'Are they *experts* who are dealing with language in their professional contexts?'

Table 5.1 Data on selected Wikipedia entries on usage items*

Title of the article	Page length in bytes	Difference in page	N of watchers	Date of page creation	Total N of edits	Total N of distinct authors	Page views April 2014	Difference to avg. page views
Average values	2,680	n/a	n/a		21.82		2,095.45	n/a
Ain't	16,040	+498.5%	58	6.9.2010	199	112	18,731	+793.89%
Double negative	25,007	+833.1%	68	9.3.2003	849	513	9,880	+371.5%
Fewer vs. Less	6,774	+152.8%	<30	17.1.2007	120	64	3,710	+77.05%
Gender neutrality	33,962	+1167.24%	111	24.8.2002	1,331	551	4,025	+92.08%
Preposition	13,486	+403.21%	<30	27.1.2006	165	85	2,033	-3.07%
Split infinitive	49,338	+1741%	118	18.11.2001	1,124	493	18,761	+795.32%
Whom	21,064	+685.97%	33	4.3.2004	400	208	2,095	-0.2%

*The data were retrieved from the information provided for each page (2014, April 20), Wikipedia article traffic statistics (<http://stats.grok.se, 2014, May 20>) and Wikimedia statistics (<http://stats.wikimedia.org/, 2014, May 25>).

Due to the fact that at least 30% of the editors remain anonymous (Viégas et al., 2007, p. 7) and many do not provide biographical data on their user websites, it was impossible to present an accountable quantitative overview of the editors' professions or involvement in language-related fields. However, by observing available user websites, it is clear that a number of editors of usage entries work in language-related professions, whereas others include language and grammar among their personal interests. An author who writes under the name of Doric Loon is Professor of Translation. On his user website, he provides information about his interests and about Wikipedia edits which greatly coincide with his real-life research interests. Another frequent contributor on the usage items pages is Daniel Freeman, who in spite of not having a job that would directly qualify him as a language expert describes himself as a user of proper English:

- (1) I am no great writer but I know the basics of proper English writing. I even use the word 'whom' correctly! When I worked at the local newspaper they called me the 'comma king' because I was an expert at knowing where commas are required, and where they are optional.

Jerry Friedman is another non-expert, but, again, a contributor who qualifies as a well-informed individual teaching physics and mathematics and providing style advice for Wikipedians on his user website, an example of which is presented in (2).

- (2) Have some doubt about 'Note that', 'Interestingly', 'It is important to note that', etc. You can often just leave them out.

Gramorak is a retired language teacher who collects early grammars of English and tries to write a grammar of the English verb. So-

SaysSunny double-majored in Astrophysics and Math, and says she cannot avoid noticing grammar mistakes and typos, which earned her the moniker Renegade Grammarian for contacting webmasters concerning grammar mistakes. Her ‘go-to’ book on English usage is Fowler’s *Modern English Usage*.

Although many contributions remain anonymous, and not enough biographical data on the contributors are available to make general conclusions, it can be observed that authors with more ‘real-life’ linguistic expertise often contribute more frequently, and assume a more relevant role in coordinating the edits than the anonymous and novice users.

5.5.2 *Wikipedia entries on usage items*

Busse and Schröder (2006, p. 71) describe usage guides as works in which the description of grammar and lexis are synthesised, contrary to the traditional division of labour between grammar books and dictionaries. Wikipedia entries analysed here do not merely describe lexical items, as does its sister dictionary project *Wiktionary*; however, they are not usage guide entries either. Instead of instructing the readers on language use, Wikipedians are urged to create informative entries.

- (3) Wikipedia is not in the business of saying how words, idioms, phrases etc., should be used (but it may be important in the context of an encyclopedia article to discuss how a word is used.⁷)

Nevertheless, the secondary Wikipedia pages, i.e. the Talk pages, demonstrate that prescriptive attitudes are present to a great extent among the contributors and attempts are made in the initial stages of

⁷ See ‘Wikipedia: Wikipedia is not a dictionary’ wiki/Wikipedia:Wikipedia_is_not_a_dictionary.

creating the respective entries to include the correctness labels related to the particular usage items, as is evident in (4).

- (4) The article reads like a style manual. The majority of the article is focused on how to use who/m, with little focus on the historical development. Look at that tiny section on its history, it doesn't even tell us what the Old English and Middle English forms were! The tone of the article has a very prescriptivist attitude, reading more like a random book on grammar than an encyclopedic article. This is unacceptable; all linguistic articles on Wikipedia are descriptivist, sensibly so. Cntrational

I agree. That is a risk with articles involving common grammatical foibles and hobbyhorses. See also *Apostrophe*, which needs to be guarded against amateurism of the less benign sort. So please: do more than your one edit so far, to improve the article. NoeticaTea?⁸

However, although they are often overtly expressed, prescriptive attitudes and negative value judgements of usage items rarely make it into the encyclopaedia itself, due to the interventions of Wikipedia entries such as the one in (4). In order to explore this phenomenon, I will here first use corpus linguistics tools in order to illustrate the differences between Wikipedia entries and usage guide entries (see §5.5.3) and, secondly, illustrate the discourses situated in the prescriptivist-descriptivist discourse on the Talk pages as well as the dynamics that prevent them from becoming embedded in the respective Wikipedia entries (cf. §5.5.4).

⁸ See 'Talk: Who (pronoun)' [http://en.wikipedia.org/wiki/Talk:Who_\(pronoun\)](http://en.wikipedia.org/wiki/Talk:Who_(pronoun)).

5.5.3 Corpus-based comparison of Wikipedia entries and usage guides

For this analysis two sub-corpora on seven usage items were created, the first being made up of Wikipedia entries, and the second from the usage guide entries available through the HUGE, in order to explore the potential differences between the two text types. The exact numbers of entries per usage item, and the number of words per each topic and per sub-corpus are available in Table 5.2.

Table 5.2 Sub-corpora on selected usage items: HUGE and Wikipedia entries

Usage Item	N of entries	N of usage guides containing an entry for the item	Total N of words per usage problem in HUGE (combined all entries)	Total N of words per entry in Wikipedia
Ain't	29	23	8,764	2,057
Double negative	49	40	15,440	3,394
Fewer vs. less	73	58	14,435	919
Gender neutrality	40	34	14,633	3,976
Preposition stranding	51	47	21,212	1,801
Split infinitive	71	59	34,121	6,473
Whom	87	60	38,840	2,850
			Total: 147,445	Total: 21,470

In order of popularity among usage guides, the usage items are listed as follows: *Whom*, *Split infinitive*, *Fewer vs. less*, *Preposition stranding*, *Double negative*, *Gender neutrality* and *Ain't*. If we compare this order to the number of views of the Wikipedia entries in Table 5.1 (*Split infinitive*, *Ain't*, *Double negative*, *Gender neutrality*, *Fewer vs. less*, *Whom* and *Preposition stranding*), we can see that there is a considerable discrepancy between what users seem to find of most interest

and what is mostly written about. For example, the most popular item in usage guides, *Whom*, is very poorly frequented on Wikipedia when compared to the other usage items. These numbers, however, cannot be compared one-to-one, since we need to take the temporal dimension into consideration, because the usage guides in HUGE date back to the late eighteenth century, and the first usage item entry in Wikipedia was created in 2001. The page view statistics on Wikipedia entries are an indicator of how interested people are in certain topics. Thus, they provide a more accurate and data-based account of the actual usage problems people are interested in, in contrast to the usage problems selected by the usage guide authors.

In comparing the two sub-corpora, the entries in HUGE and the Wikipedia entries, I used the web-based Wmatrix tool (Rayson, 2009) for finding key words, and subsequently key semantic domains that would help determine significant differences between the Wikipedia-type entries and those found in usage guides. The Wmatrix tool (Rayson, 2009) enables access not only to the traditional tools of corpus linguistics, such as key words, concordance lines and frequency lists, but also to both automatic part-of-speech and USAS semantic taggers (Rayson et al., 2004). In this analysis I focussed on the results obtained from comparing the key semantic domains via the USAS semantic tagger that enables the researcher to identify relevant semantic categories which are overrepresented in one corpus when compared to another. Since the two sub-corpora were small in size, especially the Wikipedia sub-corpus, instead of focussing on key words, I decided to focus on key semantic domains, which facilitate the recognition of small fre-

quency items as key if they belong to the same domain. In the analysis introduced below, I present the top key categories, which were calculated by using the log likelihood test. The minimum key value of the log likelihood statistical test was set to 15.13, which corresponds to $p < 0.0001$, and the minimum word frequency from each domain was set to five occurrences. The results of the analysis are shown in Tables 5.3 and 5.4. In Table 3, the lexical items overrepresented in the HUGE sub-corpus are categorised into four different semantic domains. These domains are based on the USAS tagging system, but I also slightly modified them to better fit the purposes of the analysis.

Table 5.3 Positive semantic domains in HUGE

Positive semantic domains in HUGE	Lexical items
Pronouns	<i>it, one, we, they, you, what, he, which, our, these, those, your</i>
Downtoners	
(a) Minimisers	(a) <i>hardly, scarcely, at least, barely, at all</i>
(b) Compromisers	(b) <i>quite, pretty, rather, sufficiently, reasonably</i>
Evaluation: Inaccurate	<i>incorrect, wrong, missing, error, mistake, ungrammatical, blunder</i>
In power	<i>power, rule, govern, master, hierarchy, insist, order, force, upper-class</i>

The four semantic domains over-represented in usage guides when compared to Wikipedia are *Pronouns*, *Downtoners*, *Evaluation: Inaccurate*, and *In power*.

The use of pronouns in register variation has been addressed in various corpus-based studies conducted by e.g. Biber (1995, p. 30). First- and second-person pronouns have been repeatedly found to be related to interactiveness, whereas third-person pronouns tend to be

related to the narrative dimension of register variation (Biber 1995, p. 151).

Usage guide authors guide the readers and provide advice. They are more narrative in their style than the factual Wikipedia entries, and they occasionally address the reader directly as in (5).

- (5) We do know that it had an earlier spelling an ‘t (or sometimes a ‘n’t), which you can see would not be difficult to derive phonologically from are or am; (Ward, 1989, p. 60)

According to Quirk et al. (1985, p. 452), downtoners are intensifying adjuncts that lower the effect on the force of the modified verbs. The difference between *minimisers* and *compromises* in Table 5.3 is, however, a matter of degree. Minimisers realise a greater degree of intensity and negate the full implication of the predicate. The use of downtoners and other types of hedging devices observed in (6) can be well explained in pragmatic terms.

- (6) After reading these we can perhaps conclude that the decisive influence is probably the vague impression beforehand that whom is more likely to be right; but it need hardly be said that slapdash procedure of that kind deserves no mercy when it fails. (Fowler, 1965, p. 707)

The authors are attempting to emphasise an orientation to the reader by using first and second person pronouns and hedging devices through which they attempt to gain the reader’s acceptance of the proposed arguments (cf. Hyland, 1998, pp. xiii–ix).

In a study of prescriptivist language in letters to the editor (§3), it will be demonstrated that the choice of lexis that indicates negative evaluation is typical of prescriptivism. In prescriptivist texts marked,

nonstandard forms are compared to the non-marked standard ones, and labelled *incorrect* as in (7).

- (7) A double negative does, however, sometimes survive accidentally and incorrectly in Mod.E., especially in conversation. (Treble and Vallins, 1936, p. 65)

It should be noted though, that unlike in the usage guide *A.B.C. of English Usage* (1936) in (7), newer usage guides do not directly refer to usage as ‘incorrect’ *per se*. They rather refer to older sources through which such attitudes are secondarily transmitted in the form of citations, as in the following example from Webster’s *Dictionary of English Usage*, which cites Fowler on the usage of *ain’t*.

- (8) By 1926 H. W. Fowler could view first-person use of *ain’t* quite differently from other uses: *A(i)n’t* is merely colloquial, & as used for *isn’t* is an uneducated blunder & serves no useful purpose. (Ward, 1989, p. 60)

Finally, the lexical items from the domain *In power* are statistically overused in comparison again in Wikipedia entries due to the common reference to rules in the language of usage guides. See (9) for an example of this.

- (9) The *OED* lists numerous other examples (15-20C.) of the breakdown of formal grammatical rules governing *who* and *whom*. (Fowler, 2000, p. 847)

Table 5.4 demonstrates the domains and their respective lexical items that are over-represented in Wikipedia pages when compared to the usage guide entries. Five different domains are listed, but a number of them can be combined and ascribed to the same phenomenon. Thus, numbers and the much more common references to personal names and

the media provide common evidence of one characteristic to the online encyclopaedia (see above) that is generally largely lacking in usage guides, i.e. referencing. The *Numbers* in the Wikipedia corpus most commonly refer publication dates cited and the *Personal names* are often names of authors. The domain *The Media: Newspapers* consists mostly of words that are parts of titles of journals and other types of publications.

Instead of providing *ipse dixit* judgements, such as the one common in usage guides, Wikipedia authors, guided by the Verifiability principles (see above) are obliged to support their arguments by providing external sources, which do not include only individual linguistic authorities, but also up-to-date linguistic research. External sources in Wikipedia usage item entries include usage guides, grammar books, newspaper and journal articles, encyclopaedias and academic weblogs (such as *Language Log*). Instead of referring to grammatical rules, the Wikipedia entries focus on reporting linguistic descriptions of different types of usage in different varieties of English, and, occasionally, the editors also introduce comparisons with other languages regarding a particular usage item.

For this reason, the semantic fields made up of lexical items from the field of linguistics, *Language, grammar and linguistic processes*, and *Languages and language varieties*, are over-represented when compared to the usage guides.

An example of the types of accounts that rely on the specialised terminology of linguistics can be seen in the excerpt from the entry on the *Double negative* in (10).

- (10) A similar development to a circumfix from double negation can be seen in non-Indo-European languages, too: for example, in Maltese, *kiel* ‘he ate’ is negated as *ma kielx* ‘he did not eat’, where the verb is preceded by a negative particle *ma-* ‘not’ and followed by the particle *-x*, which was originally a shortened form of *xejn* ‘nothing’ - thus, ‘he didn’t eat nothing’.

Table 5.4 Negative semantic domains in HUGE

Negative semantic domains in HUGE	Lexical items
Numbers (dates and page numbers)	<i>1925, p. 25, 1998, etc.</i>
Languages and language varieties	<i>British, standard English, Cockney, Australian, Aussie, American, Scots, French, German, Greek, Latin</i>
Personal names	<i>Eric Partridge, Jonathan Swift, Dickens, George Bernard Shaw, H. W. Fowler, Shakespeare</i>
The Media: Newspapers	<i>article, press, magazine, newspaper, correspondent, headline</i>
Language, grammar and linguistic processes	<i>utterance, token, grammar, verb, plural, denote, sign, clause, imply, proverbial, speakers, usage, vowel, pronunciation, dialects, language</i>

To return to Biber’s dimensions, which were mentioned briefly in relation to the discussion of Table 5.3, the linguistic items that are characteristic of Wikipedia entries are more representative of informative language than of what we normally find in usage guides. Precise names (*Personal names*) and references (*Numbers, The Media: Newspapers*) are given, and the specialised terminology of descriptive linguistics is used, which stands in stark contrast to the rule-oriented, guiding and often prescriptive language of usage guides.

5.5.4 Comparing Wikipedia Talk pages and entries on usage items

As previously mentioned, the reason why Wikipedia entries maintain a high level of quality and objectivity is that general consensus exists

between Wikipedians, which is for the most part obtained by the experts, i.e. senior Wikipedians who have been around for some time. In practical terms, the quality is directly guided through the application of Wikipedia principles. In this section I will show how prescriptivist attitudes have failed to enter the articles on the described usage items due to the application of Wikipedia principles.

As mentioned in the previous section, one of the ways in which the Wikipedia entries on usage which I analysed differ from the usage guide entries is referencing, which is due to Wikipedia's principle of Verifiability. The references and examples of proper usage in the prescriptivist tradition are the renowned literary figures, and the 'one's social and intellectual betters' (Landau, 1979, p. 4). In the HUGE subcorpus analysed here there are many instances which illustrate that the 'intellectual betters' are often taken as exemplary language users, as can be seen in (11).

- (11) Than them (=than they) is used by Adelaide Procter; than me, occasionally in Shakespeare, Swift, Prior, Pope, Dr. Johnson, Southey, Thackeray, Bulwer, and Clough; than him in Shakespeare, Johnson, and Kingsley; than her in Boswell and Prior. It has been said in earlier paragraphs that these phrases are found in the King James Bible, Caxton, the Genevan Bible, Goldsmith, Scott, Beddoes. (Hall, 1917, p. 293)

The discussions on Wikipedia Talk pages illustrate that entries without appropriate references are deleted from the entries (12), classical authors are not always seen as appropriate sources for illustrating acceptable usage (13), and even the traditionally reputable sources do not escape the Wikipedia editors' critical scrutiny (14). In (12), the edi-

tor called Drjamesaustin is negotiating his edit with another experienced editor. The discussion is resolved when he admits to an insufficient basis for his addition to the entry, which is based on intuition and interpretation, instead of a reputable source.

- (12) That's fine with me. I can cite no references to back up my interpretation—merely my Latin-influenced view of grammar—so will gladly bow to higher authority. Thanks Jerry! Drjamesaustin⁹

Some authors of usage guides uncritically list classical sources to illustrate good usage. In (13), the editor is questioning both the sources, and the appropriateness of comparing the usage in poetry to the everyday usage.

- (13) Likewise I don't see that the use in poetry is particularly important. We do need the example from Shakespeare and the one from Burns, but I don't think the Shakespeare one is necessarily 'to good effect'. Anyway, much though I like poetry, I'd say it's outside the mainstream of usage, so it doesn't belong in the lead. JerryFriedman

Finally, even reputable sources such as Garner's *Modern American Usage* (1998) referred to in (14), are critically evaluated and compared with other sources in order to report on insights that are most plausible or most widely accepted by the expert community. In this way, the editors avoid repeating myths that could have been presented and reported on in numerous sources but have nevertheless been disputed, such as in our example, the statement that the eighteenth-century

⁹ See 'Talk: Split infinitive' http://en.wikipedia.org/wiki/Talk:Split_infinitive.

grammarian Robert Lowth proscribed against the split infinitive (cf. Tieken, 2010).

- (14) The article now says that Lowth did NOT proscribe against the split infinitive. Every reference I have ever seen says that he did, and he is cited in Garner's latest usage manual. Where is the evidence he did NOT create this proscription? Manning

Manning, if you've want, you can add the cite from Garner to footnote 13 as more evidence that reputable people believe in the myth. JerryFriedman

Another Wikipedia principle that works in favour of the lack of prescriptive argumentation in the articles is the Neutral Point of View (NPOV) principle. In the words of Cameron (1995, p. vi), 'language is, notoriously, something which engenders strong feelings'. The expressions of attitudes towards language, although working against the NPOV principle, resurface in the Talk pages, while off-topic remarks on usage of the for-and-against type are common in Talk pages (15).

- (15) Split infinitives are absolutely never acceptable as any grammarian would tell you. The idea that they are acceptable in some contexts or that this is a controversial issue among grammarians is simply a common misconception. For a reliable source, just ask any English Professor or even any K12 English teacher. The Mysterious El Willstro

You are entitled to your usage. You'll be very lucky indeed if you can find a university teacher of English who agrees with you, but K12 teacher might just be possible. Good luck with that. Doric Loon (...)

If she'd like to contribute information with reliable published sources to the article, she's welcome to. JerryFriedman

The closed off-topic discussion in (15) entitled ‘Hear me out...’ is an example of an exchange of opinions about the acceptability of a particular usage item. Whereas the user introducing the topic, The Mysterious El Willstro, introduced arguments based on alleged opinions of language professionals, one of the most active editors of the article responds by requesting explicit references in print form, which The Mysterious El Willstro is not able to provide. Finally, after seven exchanges, the administrators closed the topic by citing one of Wikipedia’s guidelines: ‘Wikipedia is not a forum’.

5.6 Conclusion

Discussions on language use have entered the new media since the introduction of Web 2.0, one of them being the online encyclopaedia, Wikipedia. Wikipedia is primarily defined as an online collaborative community that introduced a great shift from a one-way communication pathway between expert and lay communities into a dialogue. In the language sphere, the OED would be an example of a forerunner of such collaboration between the two communities with its history of contributions from the general public.

The traditional sources for the ‘linguistically insecure’ before the online usage advice came into the scene are usage guides. Although very popular among their readership, usage guides have been criticised by linguists for their authors’ prescriptive attitudes and lack of referencing (Peters & Young, 1997, p. 318). In this study I compared the entries on selected usage items on Wikipedia and in the database of usage guides and usage problems developed at Leiden University Centre for

Linguistics, HUGE. The Wikipedia entries on usage items have proved to be on average more visited, heavily edited and they have attracted more collaboration from Wikipedia editors than the average Wikipedia article (see Table 5.1). Some of the editors participating in the writing of the entries on usage items are language experts, and, usually, the more real-life expertise they have in language-related fields, the more editing and coordination they do in the entries on usage items. A corpus-based comparison of usage guide and Wikipedia entries has shown that usage guides tend to use a more narrative and personalised style, which focuses on the rules and the stigmatisation of nonstandard usage. Wikipedia entries, on the other hand, include significantly more references to other sources and lexical items that come from the specialised terminology from linguistics.

The qualitative analysis of the secondary Talk pages has shown that the main reason why Wikipedia entries manage to obtain a level of objectivity and avoid prescriptive accounts, in spite of the many editors' comments which include evaluation and negative attitudes to the non-standard usage, is the editors' commitment to Wikipedia principles. The principle of Verifiability influences the number of references included in each entry, while the principle of the Neutral Point of View supports descriptive as opposed to the traditionally prescriptive accounts of usage guides.

Wikipedia entries on language use are a product of collaboration of many contributors as opposed to usage guides that are usually works of individual authors. Due to this fact, as well as the guidance of Wikipedia principles, balanced discussions on Talk pages and regular edits,

Wikipedia entries on language use largely reflect critical, up-to-date accounts relying primarily on linguistics as a discipline and actual usage, instead of single authorities and traditional gate keepers. The actual impact of Wikipedia and other forms of social media on usage and their popularity when compared to other sources of advice on usage is yet to be explored. Widening the research scope beyond the discourse on language use, it would be worth analysing different phenomena of expert discourse re-contextualisation on Wikipedia in fields other than linguistics and looking into the effects and potential difficulties accompanying the communication between experts and laypeople facilitated by the web-based collaborative processes.

