

JOACHIM GRZEGA

HOW ONOMASIOLOGISTS CAN HELP WITH CONTRIBUTING TO WIKIPEDIA

Abstract

In this article Wikipedia is presented as the most important everyday venue for knowledge management. The three different main styles are described: namely the ones of articles, article talk pages and user pages. Then several aspects are commented on from an onomasiologist's perspective: (1) content management on talk pages (e.g. thematic structures should be preferred over linear structures), (2) evaluation of cited sources (e.g. authors should be experts, results should have appeared in acknowledged venues, facts and opinions should be distinguished), (3) expert-layperson communication (e.g. different types of definitions including the use of examples should be used, jargon can be used if explained, contents should be structured from the general to the specific, description instead of evaluation should be used), (4) linking (including setting links to one's own article in other articles) and (5) categorizing into conceptual fields. Examples are taken from the English version of Wikipedia, but generalizable to other versions as well. The final section of the paper gives a few ideas for integrating the observations of the article into high-school and university education: In every subject students should be encouraged to practice expert-novice communication through collaborating in Wikipedia; students are offered guidelines on contributing (to) articles (e.g. concerning the creation and understanding of definitions, text structure, jargon, neutral point of view, linking and categorizing) and guidelines on contributing to talk pages (e.g. the use of an "integrative style", which aims at achieving consensus between contributors and not at having administrators decide on the content of articles).

1. Introductory Remarks

Although the 16th volume of the internationally renowned series *Handbücher zur Sprach- und Literaturwissenschaft* (Brinker et al. 2000-2001) offers a broad-ranged overview of aspects of text and discourse analysis, one specific context could not be covered as it didn't exist yet in the volume's publishing year: the Internet encyclopedia Wikipedia (<http://www.wikipedia.org>)¹. *Wikipedia* is a blend of Hawaiian *wiki* 'fast' and the English *encyclopedia* and is the name of the largest and most popular encyclopedia with free on-line access for both contributors³ and readers (<http://www.wikipedia.org>). It was initiated by Jimbo Wales and Larry Sanger on 15 January 2001 as a parallel on-line encyclopedia to Nupedia, which was to feature articles written and reviewed by academic experts. While the idea of Nupedia failed due to the slowness of the process the experts were traditionally

¹ All internet pages referred to in this article were visited on April 19, 2006.

² Studies that deal specifically with Wikipedia are listed and continually updated on <http://wikibib.pbwiki.com> and http://meta.wikimedia.org/wiki/Wiki_Research_Bibliography or <http://bibliography.wikimedia.de/index.php>.

³ The basic concept is called a wiki technology, which was developed by Ward Cunningham (Cunningham/Leuf 2001). This server-side technology allows users to produce instant updates of a web page via a web interface. All pages (except for a few administrator-protected ones) have an "edit" view with a text field containing the page's contents. The user can edit this text and submit a new version that immediately replaces the older one. The older one, in turn, is still stored on a "history" subpage. This protects a page from permanent harm. Users have the option of registering, and their contributions will then be served under their profiles. On a discussion subpage you can add comments; the structure of the discussion subpage resembles the thread of a discussion forum and is mostly used in a linear and chronological manner (although it would be possible to delete and supplement any part of this subpage as in editing the main page).

used to, Wikipedia has become more and more increasing in size and in popularity. Today there are Wikipedia versions in over 200 languages and dialects, with the English version being the largest one (with over 1,087,000 articles on 19 April 2006). More and more people find interest in contributing to Wikipedia and thus become experts who want to present, and have to think about how to present, their knowledge to a large audience. Moreover, several offspring projects have developed, such as Wiktionary, Wikibooks and Wikisources.

In this article I would like to show the various styles requested when contributing to Wikipedia, the respective hints and instructions that Wikipedia offers and a few remarks from an onomasiologist's perspective as language is the most basic tool for knowledge management⁴. Since Wikipedia means, as I have already said, that more and more people, as Wikipedia authors, become experts who want to offer their knowledge to an audience that also consists of non-experts, this kind of expert-layperson communication should also be trained at school and university—a few curricular suggestions are listed in the final section of the paper. Again, everything quoted or observed concretely refers to the English Wikipedia, but can well be generalized for every language version of Wikipedia.

2. The General Features of Wikipedia

The features of Wikipedia partly differ in the single language versions. But here are some of the Wikipedia minimum features offered to the **reader**:

- search for an article
- search for a keyword (in articles or article headlines)
- random article search
- history of article versions
- listing of articles in alphabetical order
- link to all articles, ordered according to categories, subcategories and alphabet, automatically generated if an editor has written the respective Category⁵ label into an article
- Portals (= pages intended to serve as superordinate pages for specific topic areas which normally list article links in a structured way)
- list of pages that have put a link to the article currently selected
- thematically ordered help areas where readers could ask any question they may not find answered in an article (called Reference Desks)
- possibility of getting to know something about a contributor if s/he's put up a user page
- Wikipedia-internal links within an article
- Wikipedia-external links and list of printed works within an article
- links to the corresponding article in other language versions of Wikipedia
- list of awarded articles (excellent articles and good articles)
- note within an article page that shows that the page has been elected as an excellent article (feature article) or a good article
- a printable version of each page
- a news box
- the possibility of setting up a user page

Practically, anybody from anywhere in the world can participate in any Wikipedia version. In other words, Wikipedia is built by people who may be accustomed to very different

⁴ The aspects influencing the efficiency of language as a tool for referential goals has been illustrated in a state-of-the-art article by Schnotz/Heiss (2004).

⁵ Whenever I use non-standard capitalization with *Category* etc., I refer to Wikipedia entities.

conversational rules. This means that certain conversational agreements have to be set for Wikipedia. In the context of Wikipedia—we might even speak of Wikipedia culture—there are three basic subcontexts, which also require three different styles:

- Article pages
- article discussion pages (also called Talk pages)
- User pages (including the User talk pages)

3. Remarks and Recommendations from an Onomasiologist's Perspective

Reagle (2004) states: “Wikipedia, either purposefully or accidentally, reflects many findings in the literature on how to encourage productive interdependent collaboration”. This is because the guidelines themselves are the result of collective discussions. Indeed, Wikipedia gives a lot of advice on aspects to be respected when contributing to Wikipedia. Unfortunately, they are a bit scattered over various pages: “Wikipedia:List of guidelines”, “Wikipedia:Five pillars”, “Wikipedia:Policies and guidelines”, “Wikipedia:List of policies”, “Wikipedia:Manual of Style”, “Wikipedia:Guide to writing better articles”, and “Wikipedia:User page”.

I will try to collect and epitomize the most important linguistic aspects, by beginning with the style advisable for User pages, as they are the easiest ones to master, and continuing with the styles for articles and for discussion pages.

3.1. The Style of User pages

Wikipedia administrators give users a lot of freedom to create their “public private” pages as long as they present things that are relevant to their Wikipedia work and don't misuse it as homepages and as long as they don't violate copyright rules and politeness rules. Although technically possible, it is expected that others don't edit one's user page.

Indeed, if you select a number of user pages by random choice, you will find a high degree of stylistic variation, ranging from very formal user pages illustrating the user's different experiences, interests and goals to cryptic user pages that only include the main contact data to user pages that are sometimes in part very informal and even include humor and irony. This variety is encouraged and there are no stylistic requirements that would be difficult to achieve by an author. But, in my view, it is advisable that you make clear on your user page what your fields of expertise are since, according to a study, “[a] small sample of frequent Wikipedia users said that they rely on authorship information when browsing the RecentChanges page or the history page of a specific Wikipedia article” (Viégas/Wattenberg/Dave 2004: 580).

3.2. The Style of Articles

Every Wikipedia article resembles an article from an encyclopedia, every article has a reader-friendly layout and every article looks the same, features that actually contribute to its reader-friendliness (cf. Bucher 2002: 13⁶). It seems that Internet articles have to be organized with more care than printed articles in order to achieve credibility. Wittwer/Bromme/Jucks (2004), in a contrastive study on medical information, have found out that information presented in a printed magazine, independent of the form, was rated

⁶ Bucher writes: “For most of the users the new media is full of surprise so that it is more economical [sic! instead of: economic] to meet their ‘scripts’ than to present them another surprise. The more similar a website is to this prototype the higher the usability rating by the audience in the reported test.”

significantly more credible than the same information on the Internet. Surprising result: while magazine information with illustrations were judged still more credible, the addition of hyperlinks to the Internet information had no influence on their evaluation—which stands in contrast to Morkes and Nielsen’s findings (1997). With two contradictory results, the safer side is to incorporate links.

Apart from a number of lay-out conventions, we can list the following guidelines that include the use of language and text organization. The content-focussed guidelines include these policies:

- Avoid statements that die quickly (which includes wordings such as *this year*).
- Check your facts. Include only verifiable information. Cite acknowledged sources.
- “Avoid blanket terms unless you have verified them. For example, the Montgomery County article states that of the 18 Montgomery Counties in the United States, **most** are named after Richard Montgomery. This is a blanket statement. It may very well be, but is it reliable?”⁷ and “Avoid peacock terms that show off the subject of the article without containing any real information [e.g. *an important...*, *the most influential...*]. Similarly, avoid weasel words that offer an opinion without really backing it up, and which are really used to express a non-neutral point of view [e.g. *Some people say...*, *... is widely regarded as...*].” Authors should not **say** that something is important, but **show** that something is important. And if “you wish to refer to an opinion, first make sure it is given by someone who holds some standing in that subject.”
- Neutral-Point-of-View policy: “A misunderstood notion about Wikipedia is that much contention arises from its Neutral Point of View (NPOV) policy [...]: that debates arise from this seemingly impossible requirement to remain objectively neutral. Yet, the NPOV policy is quite the opposite and instead recognizes the multitude of viewpoints and provides an epistemic stance in which they all can be recognized as instances of human knowledge—right or wrong. The NPOV policy seeks to achieve the ‘fair’ representation of all sides of the dispute such that all can feel well represented” (Reagle 2004). This particular policy, together with the requirement for indicating sources, actually better promotes transparency of academic research than selectively working mass media, which is also emphasized by Kohl/Liebert (e.g. 2004).
- No-Original-Research policy⁸

The linguistic aspects are covered by the following guidelines:

- Avoid inappropriate expressions.
- “Where varieties of English differ over a certain word or phrase, try to find an alternative that is common to both”.
- “spell out the acronym or abbreviation on the first reference (wikilinked if appropriate) and then show the acronym or abbreviation after it”.
- “Where possible, avoid using jargon. But again, consider the reader. An article entitled ‘Use of chromatic scales in early Baroque music’ is likely to be read by musicians, and so technical details and metalanguage, linked to articles explaining the metalanguage, are appropriate. [...] If any jargon is used, a brief explanation should be given the article itself.” Indeed, it seems appropriate to distinguish between articles for experts and articles for novices. But even with experts you have to reflect on whether technical terms and concepts might be different in other “schools”.
- “Use short sentences and lists.”

⁷ This quote and all others from Wikipedia [i.e. those quotes that are not supplemented by a specific source from the list of references] date from April 19, 2006.

⁸ This policy appears strange with respect to knowledge distribution. What’s the danger as long as a neutral point of view is respected and as long as the original research is also published in an academic venue?

As far as text structure is concerned, the following items are relevant:

- “An article should begin with a good definition or a clear description of the topic.” (However: what is a good definition? On this see Section 4.2).
- For the structure of an article summary style is advised: “some readers need just a quick summary (lead section); more people need a moderate amount of info (a set of multi-paragraph section); and yet others need a lot of detail (links to full-sized separate articles)”. Thus starting with a conclusion is a very reader-friendly element: Morkes/Nielsen (1997) call this the “inverted pyramid style”.
- Lead section: “It should establish significances, large implications and *why we should care*.” The first sentence: “If the subject is amenable to definition, the first sentence should give a concise, conceptually sound definition that puts the article in context.” The rest: “Then proceed with a description. The definition should be as clear to the nonspecialist as the subject matter allows. If the article is long (more than one page), the remainder of the opening paragraph should summarize it.”
- Paragraphs should contain one main idea.
- Build the web: Link articles sideways to other articles, and upwards to Categories (i.e. thematic fields that you can create by using the tag “[[Category:...]]”; they are similar to links: however, they don’t lead to other articles, but to a list of all elements provided with the same Category tag). Choose appropriate grouping techniques: categories, lists, series boxes.
- It is also recommended to avoid profanity. (We should admit, though, that profanity is a subjective notion).
- At the end: “Consider the legibility of what you are writing. Make your entry easy to read on a screen. Make judicious use of devices such as bulleted lists and bolding.”
- “Make omissions explicit.”

Morkes/Nielsen (1997) also state that meaningful subheadings should be used in web-publishing, not “clever” ones—a style that Wikipedians prefer, too. As a matter of fact, they say that a text for the web should be concise (with regard to the contents), scannable (with regard to the layout) and objective (with regard to the tone).

3.3. The Style on Discussion Pages

On various general pages readers are encouraged to respect the “Wikipetiquette”, which means:

- to sign posts on Talk pages
- to state a point, but not to prove it by spamming Wikipedia
- to avoid sweeping generalizations
- to concede a point when there is no response to it
- to admit when something has been based on intuition or taste
- to respect fellow Wikipedians’ opinions
- to be careful to avoid offending people unintentionally
- to avoid personal attacks, especially to refrain from wiki-stalking people
- to assume good faith of others and to refrain from biting newcomers
- to be prepared to apologize, to forgive and to forget
- to discourage others from being uncivil
- to stay cool when the editing gets hot
- to resolve disputes on Talk pages

If there are any problems, especially with the last two items, mediation is available if needed and asked for. If such serious conflicts occur, they are predominantly connected

with questions of content. Sometimes some editors/authors may feel that an entire article should be deleted. This can only be done by an administrator, but it is a rule that the deletion of an article is preceded by a consensus-forming process consisting of three steps and a waiting period of a week. There is also an undeletion policy that allows a deleted articles to be undeleted by any administrator. If this is controversial (or if a non-administrator wishes something undeleted) this is discussed at the Deletion review board.

All in all, these are very unusual rules of communication compared to other communicative styles in the world. In his classical work on communicative styles, Hall (1976) makes a basic distinction between “low context” communication (i.e. direct style, person-oriented, self-projection, loquacity) and “high context” communication (i.e. indirect style, status-oriented, reservation, silence). Others, such as Oetzel (1995) and Slembek (1998: 32ff.), have added a third type, namely the “integrative conflict style”, i.e. a strategy where members value group goals higher than private goals, where they ban personal conflicts and where every participant in a conversation can equally utter his or her ideas and opinions—an integration of both topic and person. The style agreed on for Wikipedia Talk pages mainly resembles this integrative conflict style. However, if a conflict can simply not be resolved, an administrator will block an article or decide on an issue.

As regards text structure, Talk pages are often continued in a linear way like many discussion forums. This has one criticizable aspect that we know from discussion forums, where “it is common that the same topics are discussed repeatedly, with long-time members complaining about newbies never reading the archives” (Wagner/Bolloju [in print]). The same can be observed in Wikipedia.

4. Onomasiological Views on Contributing to Knowledge Presentation and Management

4.1. Debating Articles: Content Management on Talk Pages

Let us briefly come back to the observation that the Talk pages are often written in a linear way instead of a thematical way. It is certainly not entirely avoidable that people simply add their question or comment on the bottom of a talk page without checking whether the topic was already dealt with before and then simply wait and see if others reply. Nevertheless, it might be advisable to encourage users to look at the rest of the talk page first and make additions where the relevant issue is already mentioned—the “search” function for a webpage, the “table of contents” feature of structured pages and the fact that users are informed of changes when the talk page is on their personal watchlists, whose versions are directly contrastable, facilitate this way of knowledge management for both writers and readers.

As I have said above, the “integrative conflict style” has been revealed to be the most successful practice for groups, or communities and it is blatantly the favored strategy in Wikipedia even if there is a voting process if consensus cannot be reached otherwise. However, such a voting process is really seen as the last possible means. On the Talk page of the article “Frank Sinatra”, for instance, a discussion on whether remarks on Sinatra’s alleged connection with organized crime should be included in a section has already been going on since 9 May 2005 when on 26 July 2005 contributor Mike suggests, “I think the proper process would be to initiate a vote for deletion for that section”. Another contributor, however, says: “a vote is not a satisfactory way of resolving these kinds of

issues because it does not work towards achieving consensus”. The vast majority of Wikipedia editors is really interested in respectful collaborative writing and willing to help out people. Thus, apart from the conversational rule “be prepared to apologize, to forgive and to forget”, there should also be a rule “thank others for their help”. This will also keep others willing to support you in the future.

4.2. The Evaluation of Cited Sources

Two quite frequent problems that I’ve detected as a source for debate on Talk pages are the confusion of facts and opinions and the value of cited sources in general. A perfect instance is the discussion on the Talk page of the article on the 19th century philologist and Sanskrit expert Max Müller. The question discussed is whether Müller was a racist (postulating a supreme Aryan race) or not. The debate is going on between an Indian and two westerners. Maybe due to cultural difference, the Indian contributor, Shivraj Singh, does not seem to understand the type of source that is required for labelling a person a racist. He writes (18 January 2006), “Majority Indians believe”, which Lukas (19 January 2006) comments with “Irrelevant. I asked you whether some **reputed historian** writes somewhere in the **scholarly literature**”; Shivraj Singh adds a bunch of other sources, among them Stephen Krapp and Ivanka Kovacevic, which Lukas (19 and 31 January 2006) comments as follows: “Who is Stephen Krapp? [...] This is not a peer-reviewed or otherwise reputable academic source” and “Kovacevic is not an expert; she is an entirely unnotable literary critic who once wrote a dissertation or something about English novels and now seems to be teaching Croatian literature somewhere”. In addition to that, Shivraj Singh, does not simply summarize the assertions in the sources he cites, but interprets them. Lukas explains: “what you write above, about Oppenheimer and DNA and whatnot [sic!], again concerns the question whether Muller [sic!] was, in hindsight, factually right; not the question of why he wrote what he wrote at the time” and “You claimed Klostermüller [i.e. one of Shivraj Singh’s sources] said that Müller was a racist; I showed you that Klostermüller does not say that Müller **was** a racist. Get it, finally? Whether Müller in fact was a racist is an entirely different question.” [all emphasis in the original]. Also on the Talk page of the article on “Frank Sinatra” we can read (comment by Aucociscokid, no date), “Most of the sources you cite are news reports of one sort or another which are well known for more often reporting innuendo rather than facts. Another way of putting is: Just ‘cause its in a newspaper or on TV, don’t make it a fact.” From an onomasiologist’s viewpoint, the safest way in all these instances is to include a phrase such as “According to Source S, A is Z” in the sentence itself instead of writing “A is Z (Source S)”, since the first rendition makes the declarative sentence rather an opinion or Source S’s perspective than a statement that also appears in source S.

4.3. Expert-Layperson Communication

Unfortunately, the Wikipedia guideline pages do not really dwell on a number of specific linguistic requirements in connection with expert-layperson communication. One fundamental fact that contributors should be aware of is that language is actually just a tool to represent the extralinguistic world. Like in traditional encyclopedias, many definitions begin with “x is Z” instead of, more appropriately, “x denotes Z”, “x refers to Z” “x is a word for Z” or “x is a designation for Z”. Even if you see “x is Z” is an abbreviation for the longer renditions, the latter may still lead to a better sensitivity for the average contributor and reader, e.g. “*dialect* is a designation for (a) a regional variety of language, (b) a non-standard variety of a language, (c) a (standard or non-standard) variety of a language”. Also of note, it is vital to keep in mind that words may be used differently: first, in

contrasting expert language and everyday language, second, in contrasting expert languages, or varieties with each other. Example: In some linguistic schools *morpheme* is used to denote ‘smallest linguistic unit carrying meaning, in others it is used to denote ‘smallest linguistic unit carrying grammatical meaning’.

Moreover, the different uses reflect the differences in “framing” the world. Different speech groups categorize the world differently. Experts should anticipate these differences between expert and lay concepts. For improving knowledge representation and the comprehension of expert concepts, experts should attempt to resort to generally understandable words and telling metaphors, good examples and relations to everyday knowledge, everyday concepts and everyday life⁹.

In the case that someone adds something to an already existing article, they should first check the definition to see whether they refer to the same concept. Such a hint, essential for collaborative writing, should be included in the guidelines, as this brief definition check seems not always be done by collaborators—and this neglect may even go unnoticed. If we have a look at the beginning section of the entry “dialect”, we read: “A dialect (from the Greek word διάλεκτος, *dialektos*) is a variety of a language used by people from a particular geographic area. [...]. The concept of dialects can be distinguished from sociolects [...], standard languages [...], jargons [...] and slang.” Although *dialect* is explicitly separated from *standard languages* in this last sentence, the second sentence after this one puts them together by saying: “Standard and non-standard dialects: A standard dialect [...] is a dialect that is supported by institutions.” This should be resolved by listing, at the beginning of an entry, the various wider and narrower definitions and then stating what the basic definition for the article should be. Again, all contributors should know what concept they are adding information to.

Concerning definitions, we can, in principle, distinguish between the following types (cf. Roelcke 2001):

| | | |
|------------------------------|---|--|
| (a) Aristotelian definitions | i.e. genus proximum plus differentia specifica | e.g. “A blend is type of word-formation that is the result of crossing two words.” |
| (b) explicatory definition | i.e. enumeration of [stereo-]typical features | e.g. “Blends are crossings of words; they are a modern type of word-formation often used for modern phenomena.” |
| (c) exemplary definition | i.e. enumeration of particularly typical examples | e.g. “Blends are, for example, <i>smog</i> (< <i>smoke</i> + <i>fog</i>) and <i>brunch</i> (< <i>breakfast</i> + <i>lunch</i>).” |
| (d) synonymic definition | i.e. giving synonyms | e.g. “Blends are also known as <i>word contaminations</i> .” |

⁹ The improvement of expert-layperson communication has been the focus of several recent works by psychologist Rainer Bromme and his team, some of which are listed in the References section. Among linguists, Gerd Antos and Sigurd Wichter have initiated a specific sub-branch of linguistics which focusses on knowledge transfer and which they therefore term *transfer science* (see the book series, initiated by Antos/Wichter in 2001 with its recent volume by Antos/Weber 2005).

| | | |
|------------------------------------|---|---|
| (e) operational/genetic definition | i.e. description of the process of how the definiendum can be produced or found out | e.g. “You create a blend by sticking the initial section of one word and the final section of another word together.” |
|------------------------------------|---|---|

The incorporation of prototype, frame and script semantic aspects in definitions may also be helpful in expert-novice communication (cf. Roelcke 2001: 61). As for definition type (c) we may especially think of prototypical members; actually, however, there is a better understanding of a category if peripheral members are included as well (provided they are marked as such). Thus, an exemplary definition of *bird* could read: “Typical examples of birds, in North America and Europe, are the robin and the sparrow; a less typical example is the penguin.” Such aspects can also be integrated in explicatory definition, e.g. “Birds lay eggs and they normally fly (also this is not a necessary feature).”

Another rule that calls for a comment is that users are encouraged to avoid jargon. But we could take another stand: it is not necessary to avoid jargon, but it is necessary to explain jargon. This stands in contrast to one of Wikipedia’s guidelines. But avoiding jargon actually means to keep the novice away from (technical) knowledge instead of making him or her familiar with it (cf. also Kalverkämper 1996, Göpferich 1998: 895f.). Furthermore, contributors must see that not only the term in question is understood, but also that other technical terms are sufficiently clear (cf. Jahr 2001: 247).

In addition to that, contributors must pay attention to another problem. Schnotz (1984: 106; 2000) found out that in the description of complex subjects authors tend to proceed from the general aspects to the details without explaining the overall frame on the general, superordinate level. This makes the formation of coherence more difficult. Besides, Jahr (2001: 249) underlines that examples should not be given exaggerated space in comparison to generalizing descriptions.

With respect to text comprehensibility in general, there have meanwhile been carried out a number of studies (cf. Göpferich [1998] for a research overview of the English-speaking and the German-speaking world, Nelson [1989] for an overview of research in America, and Kintsch [1998] for a language-independent concept of text comprehension). Nelson (1989) observes that about fifty readability formulas have been developed, but: “Current knowledge about the reading process has not had a great impact on readability formulas, which do not consider such important aspects of the text as organization connectedness, and density of content; nor do they consider the actual processing demands on readers”. And she continues: “Studies show that shortening sentences does not necessarily improve comprehension [...] and can even make reading more difficult [...]. When a sentence is divided, the connective words may be omitted and the inferencing burden increased. Substituting short words for longer, more precise words can result in a less informative text [...], thereby possibly causing the reader more difficulty in constructing meaning”. And she concludes: “Because of the complexity of the reading process, valid measures of readability may never be simple.” But it can at least be attempted to pay attention to as many aspects as possible. According to Groeben (1982), who does include connectedness in his model, a medium redundancy serves comprehension best¹⁰.

¹⁰ Here mention should be made of Groeben’s observation that maximization of text comprehensibility does not automatically lead to optimal memorization of the contents; memorization is rather increased (together with curiosity) if a medium degree of comprehensibility is reached.

4.4. Links

Wikipedia wants collaborators to build a web and therefore encourages everybody to set internal links, i.e. links to other Wikipedia articles. But is every link really necessary? Is it really necessary to link the words *journalist* and *author* when the author of a book and his profession are mentioned? Doesn't a reader know what journalists and authors are? And is it really necessary to link *wife* (the link actually leads to the entry *marriage*)? Tests will have to show whether we can sometimes even speak of “over-linking”, with negative impressions on the reader.

Some Wikipedians have specialized on setting links in articles. But another element of building the web has been overlooked so far by Wikipedia. Even if they are “link-setting specialists”, what these specialists do not do and what authors themselves are responsible for is that they should see that their article is linked in the entries that are entered as links in their article. I myself, for instance, started the Wikipedia entries for “onomasiology” and “eurolinguistics”. Within a few minutes only somebody had added links to other Wikipedia entries and “Onomasiology is a branch of lexicology” was changed into “Onomasiology is a branch of lexicology” and “Eurolinguistics deals with the languages of Europe” was changed into “Eurolinguistics deals with the languages of Europe”. However, what I had to do myself—and this is part of building a web, too—was to enter a linked remark on onomasiology in the entry “lexicology” and a linked remark on eurolinguistics in the entry “Europe”.

4.5. Categories

As I've already pointed out, a Wikipedia Category tag is similar to a link, with the difference that it doesn't lead to other articles, but to a list of all elements that have been marked with the same Category tag. The Category tags appear at the end of an article. The way categorizing is carried out also calls for some remarks. I've selected 12 random articles as a mini-corpus (in order of appearance in the selection process, all viewed on 19 April 2006):

- Jolanta Dičkutė
- 99 Precepts for Opening Hearts, Minds and Doors in the Muslim World
- Rokushko
- Norlane, Victoria
- Hagecius (crater)
- Index calculus algorithm
- Billy Strange
- Pierre Mauroy
- Julius Scriba
- List of Hungarian language radio stations
- Au, Austria
- Eugénie de Montijo

I then checked their attribution to Categories. There were 65 categorizations.

Sometimes Categories already exist when a new article is being created. In this case, articles are linked to already existing Categories in their second version on the average (including the fact that in 14 instances, a Category was attached to an article in its very first version). In other words: a categorization was carried out after 43 days, 8 hours and 59 minutes on the average. Quite fast, one might say—and indeed some Wikipedians have specialized on categorizing articles. On the other hand the slowest attributions to a

Category that had already existed at the creation of an article happened after 7 months, 5 days, 13 hours and 11 minutes (three times).

Sometimes Categories are younger than the creation of an article. Then the attribution in our corpus follows, on the average, 59 days, 4 hours and 52 minutes after the Category's creation. The fastest one was with the creation of the Category (cases of 0, 3, 4 and 7 minutes), the slowest was after 1 year, 17 days, 11 hours and 45 minutes. One entry, "Rokusho", is still not labelled with a Category at all nearly eleven months after its creation (possible Categories would have been Chemistry, Chemical compounds or Japan). In 8 instances, an article was attributed to a non-existent Category. And many entries still lack the classification into relevant existent Categories, e.g. "Jolanta Dičkutė" into the Category "Living people" or "Rokushko" into the Category "Chemical compounds"—"Rokushko", as I've already mentioned, isn't categorized into anything at all.

As a consequence, it should be more highlighted that every initiator of an article should feel responsible for assigning his or her article to already existing Categories (or, if applicable, create a new Category). This should become some sort of automatism. At best, a separate category-blank or a reminder is reserved below the editing screen of the "Edit" modus.

From an onomasiological viewpoint, it is also advisable that authors/editors do not delete superordinate Categories when they enter a Subcategory. Thus, we have the strange fact that "Billy Strange" is listed in the Subcategory "American actors", but not in the Category "Actors", while "Pat Corley" is listed only in the Category "Actors", but not in the Subcategory "American actors". After "Billy Strange" was classified into the Category "People from Long Beach, California" the Category "People from California" should not have been deleted; and adding the Category "American male singers" should not have detained editors from adding also the Categories "American singers" and "Singers". The reason is that a Category page does not automatically process the entries from its Subcategories into the actual alphabetical list of elements of the main Category. Something that the Wiki technicians might still add is some type of Subcategory-into-Category generator. But while the alphabetical list of a superordinate Category could still be supplemented by checking the Subcategories, it is even more reader-unfriendly if some items are only in the more general Category, but are not listed in a corresponding Subcategory. For instance, "Leibnitz", a city in the Austrian region of Styria, is put into the Subcategory "Cities and towns in Styria". When you click this Subcategory, however, you don't find Styria's capital, Graz. The page "Cities and towns in Styria" refer the user to two other Categories: first, "Styria", where you don't find "Graz" either, second, "Cities and towns in Austria", where you finally do find "Graz". Therefore, if you're looking for cities and towns in Styria you concomitantly have to check every single item from the larger Category "Cities and towns in Austria".

Contributors should also make sure that they create reasonable Categories: Instead of "all comedies by Shakespeare" these items should better be put under the article "Shakespeare" or a Category "Shakespeare". However, it would really be tough and endless work to create Categories for all comedies from the 17th century by English authors and the like. Here an automatic intersection finder would be helpful that would enable to find all items with the features: "Category: Comedy" \cap "Category: 17th century" \cap "Category: works by English authors".

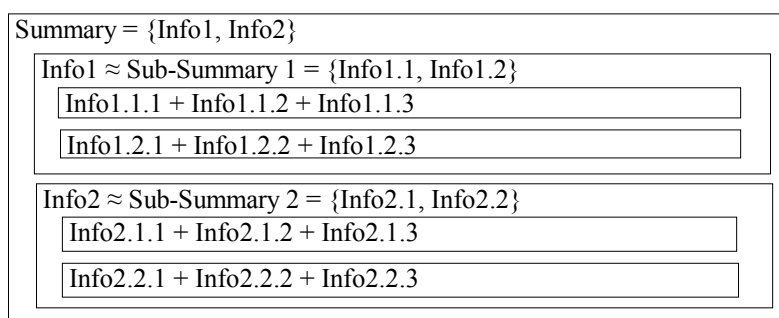
There is another problem with Categories, the problem of which Portal (or major category) to put Categories or disciplines into. Let us take linguistics as an example: Linguistics fits equally well into the History, Culture, Society, and Science portals. It seems unfortunate to put it only into one Portal (or major category). Either “linguistics” is put in various portals—or the linguistic subdisciplines (e.g. Syntax under Science, Sociolinguistics under Society and History) and so on.

5. Summary: Practising Knowledge Management in Wikipedia

Since Wikipedia becomes more and more important as a source of information, it should also play a role in education. To prepare new Wikipedia contributors they might of course be referred to articles elected as “excellent” and “good”. Yet the best way to acquire Wikipedia competence is to have learners practice writing articles themselves, to have them try out how they could best explain something they consider themselves experts for to a lay audience¹¹. If someone wants to occupy fields, i.e. a specific article, in the Wikipedia, he or she might of course start the article with an unelaborate explanation (called *stub*), but, as Viégas/Wattenberg/Dave’s study suggests, a more expanded article should quickly be worked out: “One pattern we call *first-mover advantage*. The initial text of a page tends to survive longer and tends to suffer fewer modifications than later contributions to the same page. Our hypothesis is that the first person to create a page generally sets the tone of the article on that page and, therefore, their text usually has the highest survival rate” (Viégas/Wattenberg/Dave 2004: 580f.). Therefore, a few guidelines on how to write articles may be given to the Wikipedia novice. As for the onomasiological aspects, these elements may be:

-
- (1) **Start your article your article with a definition** of the term. Remember that terms only stand for things_a, but are not the things themselves. Therefore, a term may be used in different ways by different people. If there are several definitions_b, you should include them. If applicable and possible, the definition should include an explanation of the concept a term stands for by connecting it to a superordinate term/concept and given the features that separate your concept from other concept belonging to the same superordinate one_c, (stereo)typical features_d, typical examples_e and less typical examples_f, the method for determining or creating the concept_g and synonyms_h beside the term in question. Example of a definition that encompasses all these aspects: “*Blend* is the term for_a a type of word-formation_b that is a crossing of two words; you normally get a blend by sticking an initial section of one word and a final section of another word together_g. Examples are *smog_e* (< *smoke* + *fog*), *brunch_e* (*breakfast* + *lunch*) or, less typically, *motel_f* (*motel* + *hotel*, with an “overlapping” element o). This type of word-formation is comparatively new and mostly, though not always, used with modern phenomena or events_d. Some linguists use *blend* also to include coinages such as *Monicagate*, where the second part is not the final section of a word, but an entire word_b. Synonyms are: *blending* and *contamination_h*.”
 - (2) You should structure your article in a way that you lead the reader **from the general to the specific**, so that you reach a structure like this:

¹¹ On this cf. also the didactic model *Learning by Teaching* (e.g. Grzega 2005).



This from-general-to-specific structure should be reflected in the layout, too: Article > Sections > Paragraphs > Sentences.

- (3) Don't be afraid of using jargon, but make sure that you also **explain all the technical terms and the concepts** they stand for.
- (4) Use images, bulleted lists and examples to support your explanations. But **don't write entire passages or sections on just one example**—no matter whether it is a typical or an atypical example; this would shift the focus too much away from the article term itself.
- (5) If there are several stands on an issue, **don't quote just one side, but quote all sides** (“**Neutral Point of View Policy (NPOV)**”). If you quote statements, use a phrase such as “According to X, ...”. For academic issues you should quote only studies by professionals which are articles in peer-reviewed journals, doctoral or post-doctoral dissertations or monographs from an academic book series. Don't present opinions as facts.
- (6) **Describe**, don't evaluate or judge. Don't say that something is important. **Show that something is important.**
- (7) Put a **link** on important words. At the end classify your article into relevant existent **(sub)categories** and/or create a new (sub)category.
- (8) **Incorporate your term into thematically connected articles** and put a link on your term. Only this way your article becomes connected in other articles and more people will find the way to your contribution.
- (9) Whenever you add something to an already existing article, **make sure that your additional piece of information conforms with the definition given**. If need be, supplement the already existing definition with an additional definition.
- (10) Whenever you change something in an already existing article or delete from it, **make sure that the original contribution's author doesn't feel offended**.

Comments on articles should be made on the corresponding Talk page. When you communicate on a talk page you should respect the following rules:

- (1) State a point, but do not to prove it by spamming Wikipedia. Try to reach a consensus.
- (2) Assume good faith of others. Respect fellow Wikipedians and their opinions. Be careful to avoid (unintentional) offense.
- (3) Admit when something has been based on intuition or taste.
- (4) Be prepared to apologize, to forgive and to forget.
- (5) Discourage others from being uncivil.
- (6) Thank people for help and cooperation.
- (7) Sign posts on Talk pages

In cases of emergency, you can still ask other Wikipedians, particularly the

Wikipedia administrators, for help.

.....

I would like to stress that although these ideas are presented in a linguistic journal, Wikipedia contributions and expert-layperson communication in general should be practiced not only in language classes, but in any subject—and already at high-school level, but continued into university education. Every class member could be responsible for one article and all others and the teacher can help and comment. Only this will prepare students for a society where knowledge management plays a big part in private and professional life. And the primary tool with which create and express knowledge is language.

Joachim Grzega
 Sprach- und Literaturwissenschaftliche Fakultät
 Katholische Universität Eichstätt-Ingolstadt
 85071 Eichstätt, Germany
 or:
 Englische Sprachwissenschaft
 Universität Bayreuth
 95440 Bayreuth, Germany
joachim.grzega@ku-eichstaett.de
www.grzega.de

References

- Antos, Gerd / Weber, Tilo (eds.) (2005), *Transferqualität*, [Transferwissenschaften 4], Frankfurt (Main) etc. Lang.
- Antos, Gerd / Wichter, Sigurd (eds.) (2001), *Wissenstransfer zwischen Experten und Laien: Umriss einer Transferwissenschaft*, [Transferwissenschaften 1], Frankfurt (Main) etc.: Lang.
- Brinker, Klaus et al. (eds.) (2000-2001), *Text- und Gesprächslinguistik/Linguistics of Text and Conversation*, [Handbücher zur Sprach- und Kommunikationsforschung 16], 2 vols., Berlin/New York: Walter de Gruyter.
- Bromme, Rainer / Jucks, Regina / Rambow, Riklef (2003), “Wissenskommunikation über Fächergrenzen: Ein Trainingsprogramm”, *Wirtschaftspsychologie* 3/2003: 94-102.
- Bromme, Rainer / Jucks, Regina / Rambow, Riklef (2004), “Experten-Laien-Kommunikation im Wissensmanagement”, in: Reinmann/Mandl 2004: 114-126.
- Bromme, Rainer / Jucks, Regina / Runde, Anne (2005), “Barriers and Biases in Computer-Mediated Expert-Layperson-Communication”, in: Bromme, Rainer / Hesse, Friedrich W. / Spada, Hans (eds.), *Barriers, Biases and Opportunities of Communication and Cooperation with Computers—and How They May Be Overcome*, 89-118, New York: Springer.
- Bucher, Hans-Juergen (2002), “The Power of the Audience: Interculturality, Interactivity and Trust in Internet-Communication—Theory, Research Design and Empirical Results”, in: Sudweeks, Fay / Ess, Charles (eds.), *Cultural Attitudes towards Computer and Communication*, 3-14, Murdoch: Murdoch University.
- Cunningham, Ward / Leuf, Bo (2001), *The Wiki Way: Quick Collaboration on the Web*, Boston etc.: Addison/Wesley.
- Göpferich, Susanne (1998), “Möglichkeiten der Optimierung von Fachtexten”, in: Hoffmann, Lothar et al. (eds), *Fachsprachen / Languages for Special Purposes: Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft / An International Handbook of Special Languages and Terminology Research*, [Handbücher zur Sprach- und Kommunikationswissenschaft 14], vol. 1, 888-899, Berlin/New York: Walter de Gruyter.
- Groeben, Norbert (1982), *Leserpsychologie: Textverständnis – Textverständlichkeit*, Münster: LIT.
- Grzega, Joachim (2005), “Learning by Teaching: The Didactic Model LdL in University Classes”, <http://www.ldl.de/material/berichte/uni/ldl-engl.pdf>

- Hall, Edward T. (1976), *Beyond Culture*, Garden City (N.Y.): Doubleday.
- Jahr, Silke (2001), "Adressatenspezifische Aspekte des Transfers von Wissen im wissenschaftlichen Bereich", in: Antos / Wichter 2001: 239-256.
- Kalverkämper, Hartwig (1996), "Im Zentrum der Interessen: Fachkommunikation als Leitgröße", *HERMES Journal of Linguistics* 16: 117-176.
- Kintsch, Wolfgang (1998), *Comprehension: A Paradigm for Cognition*, Cambridge: Cambridge University Press.
- Kohl, Christian / Liebert, Wolf-Andreas (2004), "Selbstorganisation der Wissenschaftsvermittlung: Quellentransparenz, Kontroversität und Qualitätssicherung in der Internet-Enzyklopädie Wikipedia", *Fachsprache* 26: 133-148.
- Morkes, John / Nielsen, Jakob (1997), "Concise, SCANNABLE and Objective: How to Write for the Web", <http://www.useit.com/papers/webwriting/writing.html>
- Nelson Spivey, Nancy (1989), "Construing Constructivism: Reading Research in the United States", <http://www.writingproject.org/downloads/csw/OP12.pdf>
- Oetzel, John G. (1995), "Intercultural Small Groups: An Effective Decision-Making Theory", in: Wiseman, Richard (ed.), *Intercultural Communication Theory*, 247-270, London / New Delhi: Thousand Oaks.
- Reagle, Joseph M., Jr. (2004), "A Case of Mutual Aid: Wikipedia, Politeness, and Perspective Taking", <http://reagle.org/joseph/2004/agree/wikip-agree.html>
- Reinmann, Gabi / Mandl, Heinz (eds.) (2004), *Psychologie des Wissensmanagements: Perspektiven, Theorien und Methoden*, Göttingen: Hogrefe.
- Roelcke, Thorsten (2001), "Was bringt die kognitive Semantik dem fachlexikalischen Transfer?", in: Antos / Wichter 2001: 51-63.
- Schnotz, Wolfgang (1984), *Textaufbau und Kohärenzbildung: Zum Einfluss der Sequenzierung von Lehrinhalten auf die Inferenzfähigkeit beim Aufbau von Wissensstrukturen*, Tübingen: Deutsches Institut für Fernstudien.
- Schnotz, Wolfgang (2000), "Das Verstehen schriftlicher Texte als Prozess", in: Brinker et al. 2000/2001: I, 497-506.
- Schnotz, Wolfgang / Heiss, Silke Frieda (2004), "Die Bedeutung der Sprache im Wissensmanagement", in: Reinmann/Mandl 2004: 41-52.
- Slembek, Edith (1998), "Grundfragen der interkulturellen Kommunikation", in: Jonach, Ingrid (ed.), *Interkulturelle Kommunikation*, [Sprache und Sprechen 34], 27-36, München/Basel: Reinhard.
- Viégas, Fernanda B. / Wattenberg, Martin / Dave, Kushal (2004), "Studying Cooperation and Conflict between Authors with *history flow* Visualizations", in: *Conference on Human Factors in Computing Systems* 2004, vol. 6.1, 575-582, <http://alumni.media.mit.edu/~fviegas/papers/history-flow.pdf>
- Wagner, Christian / Bolloju, Narasimha (in print), "Supporting Knowledge Management in Organizations with Conversational Technologies: Discussion Forums, Weblogs, and Wikis", *Journal of Database Management*.
- Wittwer, Jörg / Bromme, Rainer / Jucks, Regina (2004), "Kann man dem Internet trauen, wenn es um die Gesundheit geht? Die Glaubwürdigkeitsbeurteilung medizinischer Fachinformationen im Internet durch Laien", *Zeitschrift für Medienpsychologie* 16: 48-56.

first version received 20 May 2006
revised version received 27 June 2006