



TECHNISCHE UNIVERSITÄT DARMSTADT  
**DIGITAL HUMANITIES COOPERATION**  
UNIVERSITÄT KONSTANZ

**Marcus Müller**  
Digital discourse analysis

**LitLab**

Pamphlet #5  
**Dezember 2017**

Ed. Thomas Weitin

# Marcus Müller

## Digital discourse analysis

### Abstract (DE)

In der Diskursanalyse geht es um das Verhältnis von Sprache, Wissen und Gesellschaft. Die Rolle der Digitalen Linguistik in diesem Feld liegt darin, in Sprachkorpora, die mit Daten über Sprachverwendung und Sprecher\*innen angereichert sind, Muster zu messen, kategorisieren und kontextualisieren, die als Spuren sozialer Interaktion gedeutet werden. Dabei kann man zwei Zugänge unterscheiden: Das forensische Programm begreift Sprache als Hinweis auf Phänomene, welche ideen-, kultur- und sozialgeschichtlich relevant sind; das explorative Programm nimmt die Prinzipien der Formierung von Sprache in sozialen Kontexten in den Blick. Das Pamphlet entwickelt die methodologischen Grundlagen der Digitalen Diskursanalyse entlang der Begriffe ‚Index‘, ‚Spur‘ und ‚Kontextualisierung‘. Am Beispiel der Modalpartikelkonstruktion des Typs *doch eben, doch eigentlich* im Kontext der Bioethikdebatte stellt es die Forschungspraxis vor und plädiert schließlich für solche Forschungsmethoden, mit denen der textuelle Zusammenhang quantitativer Ergebnisse nicht aus dem Blick gerät.

### Abstract (EN)

Discourse Analysis deals with the interrelationship between language, knowledge, and society. The role of digital linguistics within this field lies in measuring, categorizing, and contextualizing patterns in language corpora enriched with data on language use and speakers. In digital discourse analysis, we interpret those patterns as traces of social interaction. We distinguish two linguistic approaches to discourse: The forensic program conceives language as indication of phenomena that are relevant to conceptual, cultural, or social history. The explorative program examines the principles of language formation in social contexts. The pamphlet draws up the methodological foundations of digital discourse analysis based on the terms ‚index‘, ‚trace‘, and ‚contextualization‘. Taking the example of the modal

particle construction of the type *doch eben, doch eigentlich* in the bioethics debate, the research practice is presented. Finally, the pamphlet advocates research methods that help not to lose sight of the textual context of quantitative results.

©2017 Marcus Müller, [mueller@linglit.tu-darmstadt.de](mailto:mueller@linglit.tu-darmstadt.de)

ISSN: 2629-7027

Ed. by Thomas Weitin

Marcus Müller

## Digital discourse analysis

### 1. What is digital discourse analysis?

Let's begin with an observation: When journalists present and moderate public discussions, their phrasing frequently contains two modal particles in succession, as seen in these examples:

*frau späht vielen dank das ist natürlich ein wichtiger appell letztlich geht es **natürlich auch** um risiken die bei der mutter entstehen*

*ja uns geht es **doch eigentlich** sehr gut*

*gibt es **denn auch** eine POSitive seite an diesem neuen gesetz für sie*

This observation may be followed up with several questions: One could ask in what syntagmatic contexts this kind of construction occurs. That would be syntactical question. One could further ask in which contexts the construction is documented, and whether it occurs more frequently with journalists and in discussions, compared to other contexts. That would be a variational-linguistic question. If you go on to ask *why* journalists speak like this in discussions, and how verbalizations like these relate to the situational, cognitive, and social conditions of journalistic language usage, you are engaging in linguistic discourse analysis.

Generally, discourse analysis studies the relationship between language, knowledge, and society. It is a flourishing field of research in the humanities and social sciences and has taken a variety of different shapes in the different disciplines. Those that do not only rely on digital or digitalized research data, but also base their analytical methods on the genuine nature of digital data, may be called digital discourse analysis.<sup>1</sup> That may mean that data is measured, and its frequency is compared with another set of data. Typically, it means that data is pre-classified on the basis of categorization models, for example by assigning strings of letters to the category 'word' (tokenization), or words to the category 'word class' (part-of-speech tagging). Tallying data alone is of course not discourse analysis, but merely an

---

<sup>1</sup>I take a datum to mean a phenomenon interpreted as a sign, which in the course of a research process is extracted from a given complexion and recontextualized. With respect to the term data and its differentiation compare Michael Bender: *Forschungsumgebungen in den Digital Humanities: Nutzerbedarf, Wissenstransfer, Textualität*. Reihe: Sprache und Wissen (SuW) 22. Berlin, Boston: de Gruyter 2016, pp. 27–37.

intermediate step. Accordingly, we need to first clarify what classification may be assumed in the connection between linguistic datum and discourse, and which methodical possibilities and restrictions arise as a result.

## 2. Datum and discourse: Language as trace of social interaction

The text bits I cited above are data that is already *interpreted linguistically*, because the form they are offered in suggests that those are communicative utterances, segmented into words. It seems however that the interpretation of those language data has not been taken very far: The citations lack punctuation and capitalization according to the rules of German orthography. The last citation on the other hand features a fully capitalized syllable. For readers with a background in linguistics, this identifies the citations as transcripts of orally produced language, with capitalization indicating stress that exceeds the expectable speech melody. The deviation from the standard notation contains another level of interpretation of the underlying linguistic data. I gave some contextual information for the citations. Thus, the linguistic data appear as a sketch of an aspect of the societal events in whose framework they were produced.

Discourse analysis deals with language *as a social phenomenon*. We are addressing a wide field of phenomena, which involves for example the variation of post-nominal attributive genitives as well as power effects of language, practices of flattery or slander, hate speech, and the linguistic perspectivation of knowledge. In other words: Anything that is part of social reality and can thus be *experienced* and correlated with linguistic patterns can become an object of discourse analysis. The experience of linguistic and social reality in itself is not discourse analysis yet, even if it occurs methodically in a controlled framework, and is subject to the transparency requirements associated with scientific research. Discourse analysts do not aim to *understand* language usage in the sense of being interested in the communicative intentions of speakers and writers. Rather, they want to *explain* language usage in context. From a discourse-analytical perspective, it is irrelevant what the journalists actually *meant* when they made the utterances with the double modal particles. The interesting question is what conditions exist when these constructs occur.

This can lead to two different lines of research in the process: Firstly, we can examine what is (probably – plausibly – typically – tendentially) the case in social reality whenever language is used in

a certain way by a certain person in a certain situation. This procedure can be compared to that of a detective securing evidence at the crime scene. He uses the traces he detected to establish a pattern, and contextualizes them to solve a murder. We can call this the *forensic* program of discourse analysis. In our example, we would simply take the evidence of the double modal particles as the basis or cause for studying the epistemic, interactional, or social situations journalists inhabit in the moment of their utterance. Secondly, one program of linguistic discourse analysis especially is explaining what forms language typically takes in given social situations, whenever language is used in a certain way by a certain person in a certain situation. We can call this type of study, which focuses on the linguistic means of expression, the *explorative* vein of discourse analysis.<sup>2</sup> Here the task would be to establish an inventory of double constructions like *doch eben*, *doch eigentlich*, their contextualization potential, and the syntagmatic patterns in which they are embedded – and then relate it to the typical situations of their usage. The study sketch introduced above will be continued below as an example of the explorative approach.<sup>3</sup>

In any case, the linguistic perspective of discourse analysis entails that we must conceive ‘language’ in the very broad sense hinted at above. This broad concept of language, however, is unsuitable for a methodically controlled language analysis. We therefore need two definitions of language, one as narrow as possible and the other as wide as possible, which are systematically correlated. The narrow definition of language S<sub>1</sub> targets the starting point of the analysis; it describes the sum of material language data configured in the discourse (i.e. written characters and speech sounds, plus possibly multi-modal signs such as images, facial expressions, gestures, proxemics) as well as the patterns of their configuration. The wider definition of language S<sub>2</sub> outlines the target area of the analysis; apart from the phenomena of linguistics of the language system, it also encompasses all aspects of social interaction that involve linguistic signs (e.g. patterns of interaction, social structures, power structures), as well as all aspects of possible thoughts, attitudes, and feelings regarding the world, which may be shared within social groups.<sup>4</sup> For discourse analysts, the intrinsic research subject is S<sub>2</sub>, but S<sub>1</sub> is a given. The task is to

---

<sup>2</sup>This distinction is introduced in relation to grammar in Marcus Müller *Diskursgrammatik*. In: Ingo H. Warnke (ed.): *Handbuch Diskurs* (Handbücher Sprachwissen) Berlin, Boston: De Gruyter (in press).

<sup>3</sup>For an example of the forensic vein see Marcus Müller & Jörn Stegmeier: *Investigating risk, uncertainty and normativity within the framework of Digital Discourse Analysis. The example of future technologies in climate change discourse*. In: Anna Olofsson & Jens O. Zinn (eds.): *Researching risk and uncertainty – methodologies, methods and research strategies*. Basingstoke: Palgrave (in press). There, we use methods of corpus linguistics to examine the ways in which risks and insecurities are made a subject of discussion in German and English media discourses. You don’t learn anything new about the German or the English language, but about the discursive formation of risk concepts in both countries. What the linguistic method yields here is a result that is interesting from a sociological perspective.

<sup>4</sup>The following thoughts in chapters 2 and 3 are a condensed version of the remarks in Marcus Müller: *Vom Wort zur Gesellschaft: Kontexte in Korpora. Ein Beitrag zur Methodologie der Korpuspragmatik*. In: Ekke-

determine tendencies and conjunctures of expression configuration in  $S_1$  with the help of open and reproducible procedures, and to then extrapolate from those tendencies and conjunctures to discourse-specific conditions in  $S_2$ , and to use the hypotheses made about  $S_2$  to obtain assertions on  $S_1$ , which in turn are testable. In other words, what Stefan Griess says on methodology of corpus linguistics is true also for discourse analysis:

... there are no meanings, no functions, no concepts in corpora – corpora are (usually text) files and all you can get out of such files is distributional (or quantitative / statistical) information.<sup>5</sup>

Now the question is how to model the connection between  $S_1$  and  $S_2$  in such a way that it offers the methodological basis for linguistic discourse research. To this end, the semiotic dimension of language is worth a closer look. According to Peirce, the fundamental relation between two phenomena that may be perceived as signs is the indexical:

An Index is a sign which refers to the Object that it denotes by virtue of being really affected by that Object.<sup>6</sup>

Peirce's well-known definition marks the point of view from which linguistic discourse analysis looks at language. The fact that the linguistic perspective is guided by the indexical aspect of linguistic signs is by no means self-evident. After all, every introductory course on linguistics teaches that linguistic signs are symbolic and owe their symbolic character to convention. What frequently is neglected in this context is that symbolic signs, provided they are natural ones, neither fall from the sky nor are codified in big language conferences, but arise whenever speakers use words in *similar* situations for *similar* communicative purposes. The recognizability and thus the symbolicity of linguistic signs arises because in the language memory of readers or listeners, a sign is "really affected" by the typical situation of its usage. The point is that every linguistic sign develops symbolic meaning in the social routines of its usage. Those are however bound up with – more or less – specific contexts. In that sense, linguistic signs ( $S_1$ ) should be understood as *traces of social situation types*.

The term 'trace' was introduced into the more recent semiotics debate most notably by Sybille

---

hard Felder, Marcus Müller & Friedemann Vogel (eds.): *Korpuspragmatik. Thematische Korpora als Basis diskurslinguistischer Analysen* 2012, pp. 33–82, here p. 34 ff., and in Marcus Müller: *Sprachliches Rollenverhalten: Korpuspragmatische Studien zu divergenten Kontextualisierungen in Mündlichkeit und Schriftlichkeit*. Berlin, Boston: De Gruyter (Sprache und Wissen) 2015, pp. 48–99.

<sup>5</sup>Stefan Th. Gries: What is Corpus Linguistics? *Language and Linguistics Compass*, 3, 2009, pp. 1225–1241. doi:10.1111/j.1749-818X.2009.00149.x, here p. 1226.

<sup>6</sup>Charles S Peirce: *Collected Papers*. Cambridge/Massachusetts: Harvard University Press 1960, p. 247.

Krämer.<sup>7</sup> It garnered a lot of attention in this debate because it connects three much-discussed dimensions of communication:<sup>8</sup>

a) Materiality

“Traces appear before one’s eyes concretely; there is no trace without a physical signature.”<sup>9</sup> Traces are always and necessarily configurations of the material in the world; they firstly refer to a physical event previous to the perceptual present, and secondly to the bio-physical process of perception itself: traces need to be seen, felt, heard, smelled. Once you describe media in communicative processes as traces, this unprejudicial materiality of communication becomes evident. And this is a central subject in the current mediality debate.

b) Semioticity

Traces only become traces when they are interpreted as such. This requires stipulations: an event that is classified as reconstructible must be presumed as cause for the trace; there must be a person willing and able to interpret a particular material configuration as trace, and an interpretation context has to exist, within which the relation between the configuration suspected of possible trace-ness and a causal event may be presumed.

c) Contextuality

Traces refer to social practices, at least in one sense, namely to those practices needed to recognize and interpret the trace as trace. If media in frameworks of communicative acts are interpreted as traces, the causative event is also shown as social practice. In addition, the term *trace* indicates that the interpretation of the material configuration as trace was not intended by the causal agent: “Traces are not made but left unwittingly. [...] And vice versa: Where something is deliberately laid and staged as a trace, it is actually not – precisely because of that intentionality.”<sup>10</sup> The theory of

---

<sup>7</sup>It was Sybille Krämer who, with reference to Carlo Ginzburg among others, rediscovered and extended the old historians’ idea of the trace for cultural studies, prominently in Sybille Krämer, Werner Kogge & Gernot Grube (eds.): *Spur. Spurenlesen als Orientierungstechnik und Wissenskunst*. Frankfurt a. M.: Suhrkamp 2007. Cf. chapters 2.4. and 2.5 in Marcus Müller, *Sprachliches Rollenverhalten* [same as footnote 4], as well as Marcus Müller: „*Symbols grow.*“ *Korpuspragmatik und Wirklichkeit*. In: Claudia Brinker von der Heyde, Nina Kalwa, Nina Maria Klug & Paul Reszke (eds.): *Eigentlichkeit. Zum Verhältnis von Sprache, Sprechern und Welt*. Berlin / Boston: De Gruyter 2015, pp. 137–157.

<sup>8</sup>The following list is taken from Müller, *Sprachliches Rollenverhalten* [same as footnote 4], p. 54 ff.

<sup>9</sup>Sybille Krämer: *Was also ist eine Spur? Und worin besteht ihre epistemologische Rolle? Eine Bestandsaufnahme*. In: Sybille Krämer, Werner Kogge & Gernot Grube: *Spur*. [same as footnote 7], pp. 11–36, here p. 15.

<sup>10</sup>*Ibid.*, p. 16, same as Werner Kogge: *Spurenlesen als epistemologischer Leitbegriff. Das Beispiel Molekularbio-*



contextualization called attention to the fact that where signs are produced intentionally, unintended side effects are always produced at the same time, and they in turn play an important part in interpretation processes, and become meaningful interactively in conversation and discourse (see below).

A separation of the observation level from the description level of linguistic research can only be simulated with methodological intent – the researcher will always remain enmeshed in his or her object. The approach outlined below – understanding contextualization processes and reconstructing them systematically with the tools of corpus linguistics – should be seen as an attempt to acknowledge **the researcher’s involvement in his or her object**, and to respond to this reconstructing activities of understanding in a methodically controlled way. Being a reader or listener, the researcher constitutes the linguistic data to be analyzed as means of communication, namely by applying precisely the methods of contextualization he intends to describe. Thus, we need a procedure that allows **the researcher to clarify for him- or herself** and others the generalizability of his or her reconstruction. In this sense, the statistical linguistic methods described below give the researcher access to an *augmented reality* of communication, where they can *measure* linguistic patterns and thus *gauge* the seriality of contextualization offers in social contexts.

### 3. Contextualization

This view was developed within contextualization theory, which is actually a network of theories.<sup>11</sup> The term ‘contextualization’ here describes methods that are used to place utterances in a context constituted by the interacting parties themselves in the course of their interactional practice. Contextualization means establishing a connection between an “empirically given (observable) datum, which the contextualizing participant selects from a repertoire of verbal and non-verbal signs”, and a “component of background knowledge”.<sup>12</sup> Following Gumperz, the indexically operative datum is called

---

*logie*. In: Sibylle Krämer, Werner Kogge & Gernot Grube (eds.), *Spur*. [same as footnote 7], pp. 182–221, here p. 118.

<sup>11</sup>‘Contextualization’ has become highly important in several areas of empirical linguistics. Contextualization theory stresses that contexts are dynamic constructs, which are constituted and constantly modeled anew by the participants in linguistic interaction. The aspects of context that become relevant in this view are primarily those which guarantee the smooth flow of linguistic interaction, namely the assessments of partner, situation, and topic. This model was proposed in John Gumperz’ interactional sociolinguistics, and in Germany was picked up first by Peter Auer in the framework of conversation analysis [bibliographical references in footnotes 12 and 13].

<sup>12</sup>Peter Auer: *Kontextualisierung*. In: *Studium Linguistik* 19, 1986, pp. 22–47, here p. 24.

“contextualisation cue”.<sup>13</sup> The essential assumption is that background knowledge is organized in the form of schemata.<sup>14</sup> Auer distinguishes schemata on five different levels:

the general schema of focused interacting, the schemata of turn-taking (i.e. recipient, listener, speaker, addressee), activity schemata, thematic schemata, as well as relationship schemata. The respective levels correlate with the following interactive problems: 1) Are we currently talking to each other? 2) Who is (currently) speaking to whom? 3) What are we (currently) doing? 4) What are we (currently) talking about? 5) What is our (current) relation?<sup>15</sup>

Three aspects are singled out separately: Firstly, Auer stresses that the framing questions he postulates need to be constantly answered anew by the participants during a conversation; potentially involving all perception data available to them. Secondly, he states that the schemata are at least partly interdependent, thus for example answering the question of the mode of interaction is a prerequisite for answering the question of the relationship between the interacting parties. Thirdly, he makes clear that interpretation schemata and interaction process are perpetually reciprocally interrelated. Thus, a new topic may bring about a reinterpretation of the mode of interaction by the collocutors, and along with it constitute a new relationship between the participants. Auer expands this model in a subsequent text: Apart from the pattern of interaction (“context brought about”), he also addresses the social-structural dimension of the interaction (esp. social roles) and the physical aspects of context given in the interaction, such as time, physical environment etc. (context brought along).<sup>16</sup>

A second version of the contextualization idea can be found in discourse analysis and understands context solely as a cognitive category. Especially the most recent work of van Dijk should be named here. Van Dijk defines ‘context’ as “the structured set of all properties of a social situation that are possibly relevant for the production, structures, interpretation and functions of text and talk”.<sup>17</sup> Here, the interface between discourse and context is cognition, just as it was for Auer – contexts do not become constitutive of discourse as objective, stable givens, but as mental “context models”.<sup>18</sup> As

---

<sup>13</sup>John J. Gumperz: *Discourse strategies*. Cambridge: University Press 1982. The German term is ‘Kontextualisierungshinweis’.

<sup>14</sup>Peter Auer: *Kontextualisierung*. [same as footnote 12], p. 24.

<sup>15</sup>Ibid., p. 27.

<sup>16</sup>Peter Auer: *Introduction: John Gumperz’ approach to contextualisation*. In: Peter Auer & Aldo Di Luzio (eds.), *Contextualization of language*, Amsterdam: John Benjamins 1992, pp. 1–37. The distinction between “brought along” vs. “brought about” originates in Volker Hinnenkamp: *Foreigner talk, code-switching and the concept of trouble*. In: Karlfried Knapp, Werner Enninger, & Annelie Knapp-Potthoff (eds.): *Analyzing Intercultural Communication*, Berlin, New York: de Gruyter 1987, pp. 137–181, here p. 143: “I would like to introduce here the dualistic concept of taking into account what is locally *brought about* through joint effort against and complementary to what is *brought along* in terms of their emergent and prestructured groundedness into the actual encounter.”

<sup>17</sup>Teun A. Van Dijk: *Ideology. A multidisciplinary approach*. London: Routledge 1998, p. 211.

<sup>18</sup>Ibid., p. 212.

leading knowledge configurations, these are systematically described on all relevant levels:

Such context models are stored in episodic memory, just like the event models are used to represent what a discourse is about. Context models, thus, represent how participants in a communicative event see, interpret and mentally represent the properties of the social situation that are now relevant for them.<sup>19</sup>

Since van Dijk propagates a discourse analysis that is interested in the social world, the concept of purely subjective mental context models cannot suffice. Therefore, van Dijk introduces the ‘social cognition’ model, which is known from social psychology:

It is important to stress that even unique, subjective models of specific events are not entirely personal. They also have important social, intersubjective dimensions. Because of earlier interaction and communication, and more generally due to their socialisation, language users have acquired various kinds of shared knowledge and other beliefs. After generalisation and abstraction, such shared general and social beliefs influence the construction of new models whose intersubjective dimensions enable interaction and mutual understanding in the first place.<sup>20</sup>

Thirdly, the notion of contextualization was functionalized in the tradition of structural language analyses. It was primarily Helmuth Feilke who developed the concept and embedded it in his theory of sociality as the locus of imprinting linguistic patterns:<sup>21</sup> The significance of linguistic signs should be seen in the ability to index ongoing strings of expression, gained by iteration following social rules. Here context is first and foremost communicatively shaped *cotext*. This structurally inspired notion of context is especially important for corpus linguistics.

For a methodology of digital discourse analysis, we need to methodically correlate the structural notion of context in the sense of ›cotext‹ and the socio-linguistic notion of context in the broader sense of ›self-reflexively available, socio-communicative orientational knowledge‹. Elsewhere I formulated an integrative working concept of context as the theory-driven foundation of a methodology of digital discourse analysis. I will briefly describe this working concept here.<sup>22</sup> Insofar as it was developed with a methodological purpose, it is designed perspectively and selectively. **A context is thus the environment of a linguistic focus construction that is relevant for (its) interpretation.**

---

<sup>19</sup>Ebd.

<sup>20</sup>Ibid. p. 6.

<sup>21</sup>Helmuth Feilke: *Sprache als soziale Gestalt. Ausdruck, Prägung und die Ordnung der sprachlichen Typik.* Frankfurt a. M.: Suhrkamp 1996.

<sup>22</sup>Marcus Müller: *Kontexte in Corpora* [same as footnote 4].

By focus construction I mean a form and function pairing as understood by construction grammar<sup>23</sup>, which indexes typical correlations of its usage. A focus construction may consist of a linguistic term, a lexically fully or partly specific phrase (*What a bummer!*; *What a x!*), or a lexically unspecific grammatical pattern (*WP DT NN!*). For the communication partners, the indexicality of the focus construction in terms of typical contexts arises from their habituated experience of language usage, while for the contextualization researcher, it arises from analysing correlations between context types and focus constructions.

The inner layer of a context consists of the linguistic cotext (including phenomena of its performance, mediality, prosody, or typography) ( $S_1$ ). The configuration of this inner context layer does not only provide clues for the interpretation of the focus construction, but also indicates its own exemplariness regarding patterns of text and interaction. Those text and interaction patterns, in turn, refer to the typology of the situations in which they are used. Notably, this means aspects such as communication's temporal frame, closeness or distance of the communication partners<sup>24</sup>, patterns of personal constellation in space, as well as typical physical environments. The situational constellation of the interaction participants and their behavior indicate their social role. In the overall picture, the exemplariness of constructions, cotexts, situations, and social role constellations provides clues for their embeddedness in thematic contexts. From this outer context layer, conclusions about deep semantic figures or epistemes may be drawn, related to groups or epochs. The totality of all context layers thus refers to the wider notion of language  $S_2$  described above. This results in an onion skin model of contextualization (fig. 1).<sup>25</sup> All mentioned contexts are constituted as and via discourse relations. In our example, the focus construction MP MP (double modal particles) indexes the activity type 'moderating', the situational context 'media discussion', and the social speaker role 'journalist' – provided it occurs in a specific configuration and a specific syntactic cotext (see below).

Locally, contextualization cues are conducive to comprehension, while structurally, they serve

---

<sup>23</sup>For more on construction grammar, see Kerstin Fischer & Anatol Stefanowitsch (eds.): *Konstruktionsgrammatik. Von der Anwendung zur Theorie*. Tübingen: Stauffenburg 2007.

<sup>24</sup>For more on this concept, see Peter Koch & Wulf Oesterreicher: *Sprache der Nähe – Sprache der Distanz. Mündlichkeit und Schriftlichkeit im Spannungsfeld von Sprachtheorie und Sprachgeschichte*. In: Romanistisches Jahrbuch 36, 1986, pp. 15–43.

<sup>25</sup>Please note that the onion skin model of contextualization is not an attempt to describe the psychological reality of the cognitive process. A psychologically realistic model would have to forgo the layers, which suggest a hierarchization of the different dimensions of context. Rather, the different dimensions would have to be modeled as coequal aspects of the cognitive process, which are mutually dependent and constantly influence each other reciprocally. Moreover, the direction of contextualization suggested here, which indicates a bottom-up process, would have to be complemented by arrows pointing in the other direction as markers of a top-down process. The onion skin model is deliberately selective and reductionist, in order to elucidate which aspects of the contextualization process may be operationalized for a methodology of corpus-linguistic contextualization research.

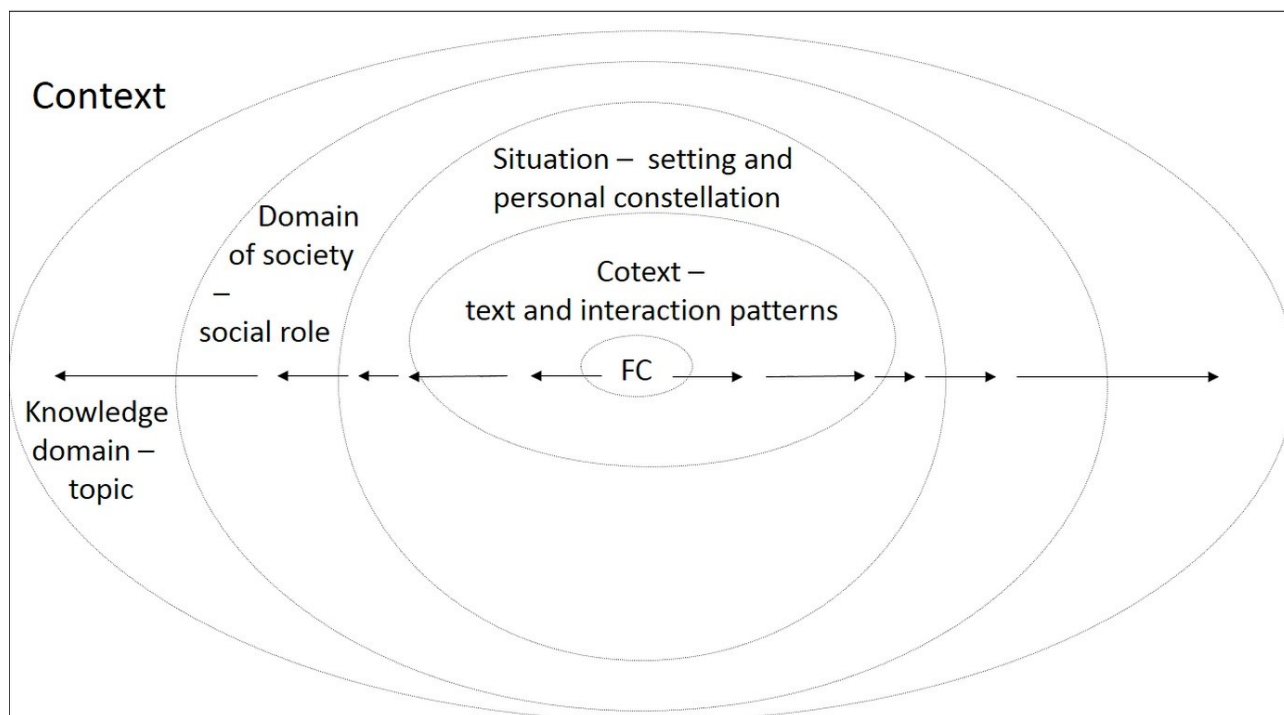


Abbildung 1: Fig. 1: The onion skin model of contextualization – FC = focus construction.

Taken from Marcus Müller: *Kontexte in Corpora* [same as footnote 4], p. 50.

to construct and organize the social symbolic universe. Retracing series of such contextualization processes, we can reconstruct the sociocognitive order of discourse. In so doing, we need to distinguish between first-order (emic) and second-order (etic) contextualizations: <sup>26</sup>

emic: intra-communicative contextualization by the communication partners

intuitive heuristics

perception-based

dynamic context models

„online“<sup>27</sup>

etic: systematizing reproduction by the linguist

forensic heuristics

corpus-based

<sup>26</sup>The conceptual pair ‘emic’ and ‘etic’ is used here in the vein of linguistic anthropology and the empirical social sciences. For the discussion of the concepts, see the contributions in Thomas N. Headland, Kenneth Pike & Marvin Harris (eds.) (1990): *Emics and Etics. The Insider/Outsider Debate*. Newbury Park et. al.: SAGE 1990. They refer to the perspective of category formation in the research process: ‘Emic’ thus means ›from the perspective of the participants of a social structure or an interaction‹, while ‘etic’ means ›from the perspective and conceptual logic of the researcher‹.

<sup>27</sup>Peter Auer: *Online-Syntax – oder: was es bedeuten könnte, die Zeitlichkeit der mündlichen Sprache ernst zu nehmen*. In: *Sprache und Literatur* 85, 2000, pp. 43–56.

static context parameters

reconstructive

We need to keep in mind that corpus linguistics always deals with serializations of language, while supposing and comprehending are capacities of the individual. Drawing conclusions from a general result to an individual one is inadmissible. Therefore, corpus pragmatics has two options: Either it interprets its semiotic results, which generally pertain to frequencies of a focus construction relative to a usage context, as probabilistic framework of meaning and understanding in the respective context addressed. Or it uses the serial results to support interpretations of individual processes of supposition and comprehension in qualitative analyses. The concept of contextualization understood in this way is thus meant to legitimize a method of linguistic forensics, with which to draw conclusions from cotexts trackable by corpus linguistics to socio-pragmatic contexts.

Discourse linguistics is familiar with the metaphor of discourses as stretched-out conversations.<sup>28</sup> One example would be the term ‘societal conversation’ as a synonym for discourse.<sup>29</sup> Conversational analysis has brought forth an important methodical guiding maxim to contextualisation analysis: Analysts need to demonstrate that linguistic phenomena are indexed as contextualization cues by the conversation partners themselves. This is done by sequential analysis. However, only corpus linguistics is capable of transferring this maxim to discourse analysis in a controlled manner. This can be done by making plausible that specific functional aspects of those phenomena are offered, affirmed and ratified as contextualization cues by the agents in discourse. Therefore, it has to be proven that a focus construction is serialized within a given context type. This results in the following methodological principles for the discourse-analytical study of contextualization:

- a) a) Meaning and understanding cannot be examined with corpus-linguistic methods.

Emic or first-order contextualization processes are individual and dynamic psychological processes. They cannot be directly studied with corpus-linguistic methods. Experimental methods or participatory observation are suited for their study. However, etic or second-order contextualization processes can be subjected to corpus analysis. They are based on analyses of correlation between linguistic pat-

---

<sup>28</sup>Namely in the wake of Ehlich’s concept of a ‘stretched-out communication situation’; cf. Konrad Ehlich: *Zum Textbegriff*. In Anneli Rothkegel & Barbara Sandig (eds.), *Text – Textsorten – Semantik. Linguistische Modelle und maschinelle Verfahren*. Hamburg: HBV 1984, pp. 9–25.

<sup>29</sup>Cf. Sven Kersten Roth: *Interpersonale Diskursrealisationen – Überlegungen zu ihrer Integration in die diskurssemantische Forschung*. In Ingo Warnke & Jürgen Spitzmüller (eds.), *Methoden der Diskurslinguistik. Sprachwissenschaftliche Zugänge zur transtextuellen Ebene*, Berlin, New York: de Gruyter 2008, pp. 323–358.

terns and typical contexts of their usage.

b) b) Digital contextualization research must yield viable results.

If a linguistic phenomenon is addressed as a contextualization cue, this must have a justification in the linguistic data that was analyzed. If a linguistic phenomenon (focus construction) appears with significant frequency in a specified context of usage, the justification is considered as given. A second-order contextualization cue determined in this manner must not simply be issued as a first-order cue. It may however count as interpretive background for the qualitative interpretation of text and conversation.

The study of contextualization in this sense needs corpora, whose linguistic data are marked with metadata for the aforementioned contextual levels ‘situation’, ‘social speaker role’, and ‘thematic discourse’, or at least those that were assembled with regard to a contextual dimension. Understood in this way, the study of contextualization is always based on a comparison of the linguistic conditions in two or more corpora.

## 4. Corpus linguistics and discourse analysis

The observation I quoted at the beginning of this text is taken from the context of a study on linguistic role behavior in the bioethics debate.<sup>30</sup> We used this topic for our research, compiling a corpus from contemporary utterances on the bioethics debate. It collects transcribed oral utterances (genres: discussion, interview, statement, call-in, poll), written utterances (genres: internet forum, policy paper, written debate contribution, reader’s comments), and verbatim records (genres: lecture/talk, hearing, parliamentary debate, TV debate). At the time of the analyses here presented, the corpus comprised approx. 3.8 million word forms.<sup>31</sup>

---

<sup>30</sup>Marcus Müller: *Sprachliches Rollenverhalten* [same as footnote 4].

<sup>31</sup>These utterances are oral and written public contributions to the bioethics debate in the German language from the years 2000–2010. The corpus is available in XML format in CQPWeb@discourselab. In addition to the specific information on the situation, the context layers social speaker role, thematic context, class, and medium are stored as metadata for each individual utterance (print or electronic text, blog post, oral contribution from a discussion), allowing for the performance of correlation studies for these context aspects. The thematic contexts included are ‘grüne Gentechnik’ (green genetic engineering), ‘allgemeine Bioethik’ (general bioethics), ‘Sterbehilfe’ (assisted suicide), ‘Lebensbeginn’ (beginning of life), ‘rote Gentechnik’ (red genetic engineering), ‘Stammzellenforschung’ (stem cell research), ‘Transplantationsmedizin’ (transplant medicine), ‘Reproduktionsmedizin/PID’ (reproductive medicine/PGD). A detailed corpus description may be found in Müller, *Sprachliches Rollenverhalten* [same as footnote 4] pp. 125–149. The individual data sources of the corpus texts are listed under <https://discourselab.de/resources/#heideko5>.

The research question for the study the records quoted at the beginning of this text are taken from was this: Are there speaker-role specific usage patterns of modal particles in the discourse on bioethics? Modal particles are non-flectional modifiers of utterances such as *denn*, *doch*, *eben*, *ja*, *halt*, *mal*, or *wohl*. They have homonyms in other parts of speech, and they always occur in an unstressed position. Modal particles mark speakers' attitudes towards propositions with a view to interaction partners, thus they index

- partner and role hypotheses,
- situation hypotheses,
- speaker hypotheses about shared knowledge relevant for comprehension, and
- speaker hypotheses about embedding utterances in discourses.

The problem inherent in this method is that modal particles are not automatically trackable in corpora, because they cannot be classified as such by part-of-speech taggers. Like many other German-language corpora, our corpus is tagged with the Stuttgart-Tübingen tagset.<sup>32</sup> Within this tagset, modal particles fall under the ADV class, i.e. “adverb.” ADV is some sort of leftover category for non-flectional words that are neither adpositions, conjunctions, or easily identifiable particles (e.g. negation particles). To annotate modal particles, I proceeded step by step: search with a list of the primary class of verbalisms which research describes as modal particles and TreeTagger classifies as ADV: *denn*, *doch*, *eben*, *halt*, *ja*, *mal*, *wohl*.<sup>33</sup> The sorted concordances were preadjusted on the basis of the syntactic position features of modal particles, and finally annotated manually. On that basis the distribution of modal particles rendered in figure 2 emerged, differentiated according to social speaker roles and medium.

The following conclusions can be drawn from the distribution analysis: The medial context domain of modal particles is, as research has frequently asserted, orality. The usage domains with a view to the social speaker roles in oral interaction are ‘journalist’ and ‘layperson’. The social role domain

---

<sup>32</sup>Cf. Anne Schiller, Simone Teufel, Christine Stöckert & Christine Thielen: *Guidelines für das Tagging deutscher Textcorpora mit STTS*. 1999. Online at <http://www.sfs.uni-tuebingen.de/resources/stts-1999.pdf> [last accessed on Nov. 12, 2017]. The coarseness of the STTS is a particular problem when tagging oral language data. For more on this issue, and for a development approach to tagging dialogue particles, cf. Swantje Westpfahl & Thomas Schmidt: *POS für(s) FOLK – Part of Speech Tagging des Forschungs- und Lehrkorpus Gesprochenes Deutsch*. In: JLCL, Band 28 (1) 2013, pp. 139–153.

<sup>33</sup>The modal particle *schon* is missing, since a sufficiently intersubjective stable disambiguation of readings in test runs with several annotators could not be guaranteed. The other modal particles can be defined unambiguously on the basis of a catalog of features.



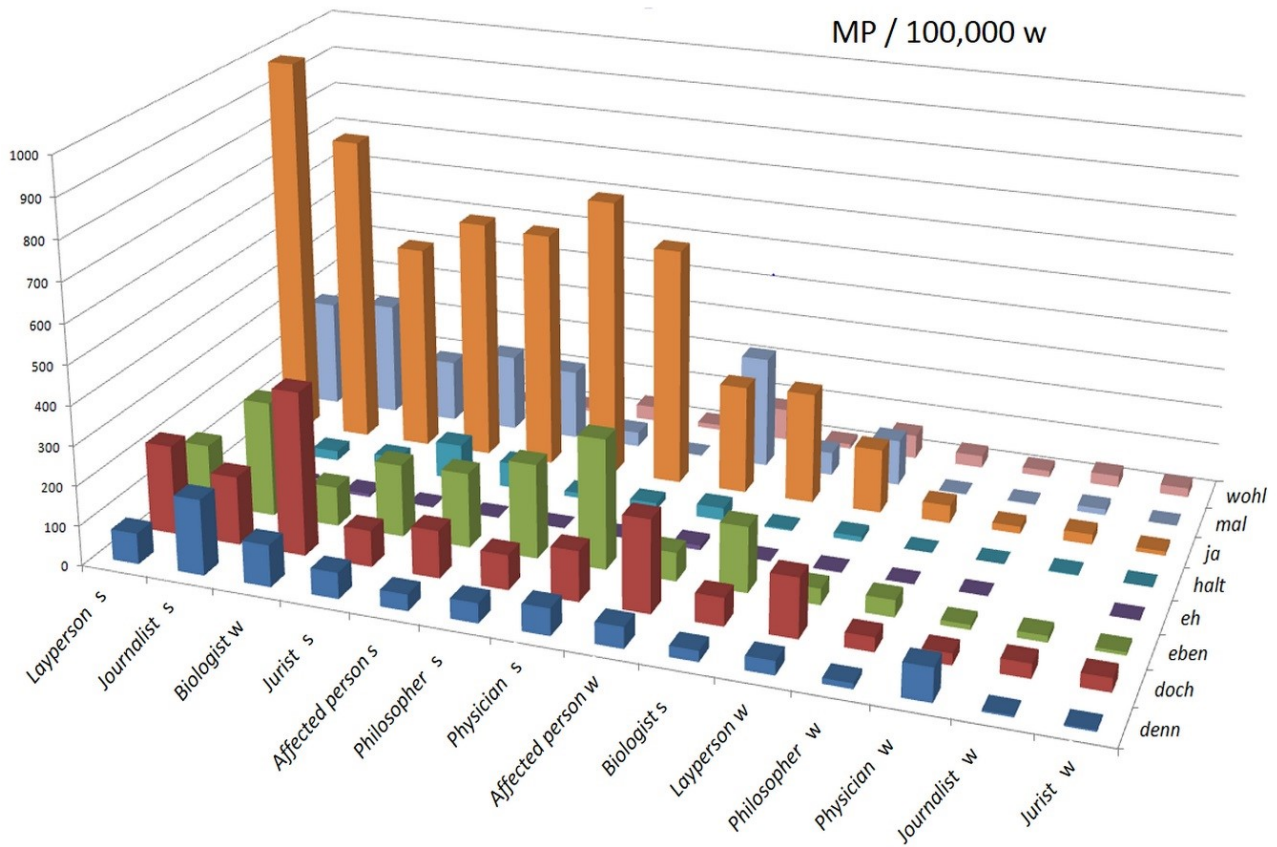


Abbildung 2: Fig. 2: Distribution of modal particles in the bioethics corpus relative to social speaker roles and medium; s = spoken language, w = written language, MP = modal particle.

in written interaction is that of biologist. The latter is easily explained by the fact that biologists in the corpus are primarily involved in communication in forums on green genetic engineering, where the strongly interactive context and the agonal mode of interaction lead to a relatively frequent use of modal particles. There is a notable negative correlation between the modal particles *doch* and *eben*. Typically, *doch* is used when you want to back up the claim to validity of a statement from a discourse position experienced as weaker or impotent – here the usage domain is the biologist’s role in forum communication in the thematic context green genetic engineering:

*Jetzt beginnen sie doch mal selber weiterzudenken . . .*

*Nehmen sie doch Wissenschaft endlich Ernst!*

*Machen sie doch endlich mal die Augen auf!*

The role domain of *eben* on the other hand is the physician’s role in oral utterances in the media. You use the word *eben* to affirm the claim to truth connected to a statement, by referring, as if in passing, to an implied common sense that is not made explicit:

*. . . ich kann eben nicht nur pränatal während der Schwangerschaft therapieren . . .*

*. . . das noch Wichtigere ist eben die Beratung . . .*

*. . . die Prognoseeinschätzung ist eben manchmal gar nicht so einfach . . .*

The modal particle *eben* thus indexes an air of the expert, which is often assumed by representatives of the physician’s role in the public sphere with reference to the practical relevance of that expertise.

From a systematic perspective the most interesting one is the journalist’s role in orality – firstly because it exhibits the most even distribution of modal particles, and secondly because compared to the overall distribution, there is a significant frequency of double occurrences of modal particles. The quoted records from the beginning of this text were taken from this context. I now want to offer a short conversation sequence and discuss it briefly. It is part of a call-in format, and the sequence follows immediately after an overlapping sequence with a caller who was just put through:

Moderatorin: *jammern sie nicht auf hohem nivEAU(.) uns geht es **doch eigentlich** sehr GUT(.) sie haben vorhin gesagt wir haben keine wartezeiten (.) schauen wir nach großbritAnnien (.) WIR gu-*

cken NICHT ob man einer fünfundachtzigjährigen noch ein Hüftgelenk einbauen kann oder nicht

Experte: herr weber hat ja durchaus EINgeräumt dass er privat versichert ist aber wir müssen einfach FESTstellen (.) in der bundesrepublik sind fast neunzig prozent eben NICHT privat versichert<sup>34</sup>

The female presenter/moderator takes up the caller's objection and repeats it for the expert in the studio. The modal particle construction we are interested in here occurs in the second utterance segment of her contribution. The modal particles *doch eigentlich* fulfill a whole range of functions at this point: First they serve as an anaphoric index, by managing the reference back to the previous utterance unit: The utterance of modal particles only makes sense if it refers back to a previous utterance. Second the particles also function as a cataphoric index, by opening up the frame of expectation for the ratification of the utterances' common ground: By using the particles, the speaker makes it clear that she expects the conversation partner to share the claim of validity of the utterance made. Both through its back-linking and its forward-linking indexicality, the particle construction takes on an important role in building coherence in conversation. Third, the modal particle *doch* pushes the stridency level of the conversation, by marking the presented utterance as an opposing position and thereby navigating the conversation into a confrontational mode of interaction. At the same time and fourth, through the use of *eigentlich*, the utterance is attenuated and elevated into the realm of the fundamental: The moderator steers the confrontational element, which is marked by *doch*, away from the interpersonal dimension of the conversation, towards that of content. With this, she performs prospective face work: She prevents the conversation partner from interpreting the modalized utterance as a personal attack and subsequently potentially jeopardizing his role as an expert in this public situation. Furthermore, the construction conduces to the conceptual orientation of the utterance, by fifth marking it as counter-thesis (with *doch*) and sixth locating it within common ground (with *eigentlich*): On the one hand it is thus implied that the statement "*es geht uns sehr gut*" is consensus, while at the other hand it is exposed as the opposite position.

An obvious argument against this analysis would be that the particles should be analyzed separately and that there is no need to regard them as a syntactic unit. What speaks against this view however is the observation that this particle doubling is exemplary in utterances of journalists in

---

<sup>34</sup>Radio program on hospitals, *Krankenhäuser in Not, Fühlen Sie sich gut versorgt?* Broadcast on station BR2, August 19, 2008.

confrontational media discussions. The pattern is precisely that in each case, the first modal particle has an escalating function and the second one has a mitigating one. Since this pattern has a clear contextualization potential with regard to speaker role and type of situation, we can label the double particles a construction in line with construction grammar. Elsewhere I have called it the ‘*walking on eggshells*’ construction, because it marks the communicative task journalists must perform while moderating discussions, making sure that on the one hand a conversation is sufficiently dynamic, and positions are sufficiently contrasted, but on the other hand allowing all conversation partners to save face.<sup>35</sup> More examples for this construction would be:

*dem kann man **doch eigentlich** nicht kommen und sagen du musst . . . .*

*dann müssen wir uns **doch eigentlich** selbst die frage stellen wollten und könnten wir . . . .*

*dann müssen sie **doch eigentlich** die debatte und die ernsthafte debatte die . . .*

*letztlich geht es **natürlich auch** um risiken . . .*

*gibt es **denn auch** eine POSitive seite an diesem neuen gesetz für sie. . .*

*dass höhere tabaksteuer dann **eben auch** fließt. . .*

*die frage ist **doch eigentlich** gar nicht so falsch gestellt ob das nicht . . .*

The corpus also contains expansions of the construction, where both the mitigating and the escalating elements are repeated:

*aber wir können **dann eben halt doch** hoffen dass*

*das war meines erinnerns im reichsstrafgesetzbuch von achtzehnhunderteinundsiebzig **eben dann doch eben schon** biologistisch festgelegt*

As mentioned above, the medium domain of the construction is orality and its role domain is ‘journalist’. It should be noted that the allocation to the journalist’s role is probabilistic. Obviously, there are also instances of this construction produced by speakers in other roles. There is no thematic contextualization potential within the themes and topics studied in the corpus, and there is no evidence that would suggest that this construction is typical for the themes and topics of the bioethics debate, even if that cannot be illustrated here. There is however a discourse-structural feature of the bioethics

---

<sup>35</sup>Marcus Müller: *Geisterkonstruktionen. Zum Beispiel PPER ADV ADV*. In Alexander Ziem / Alexander Lasch (eds.): *Konstruktionsgrammatik IV*. Tübingen: Narr 2015, pp. 203–221.

debate which triggers the construction, namely its agonal basic structure: We can clearly recognize media discussions in which divergent positions are debated as the situational domain. The dominant speech act context is the moderation of situations like that. One aspect of this speech act important here is the task of managing the multi-addressing of utterances in media discussions. This means that contributions to a conversation in the media do not only address the participants of the conversation, but also always the recipients of the media format. Thus, the moderating person ends up with the task of working out opposing positions and keeping conversations interesting on the one hand, and making sure that tempers are kept, manners are retained, and nobody loses face on the other hand. Against this backdrop, the communicative achievement of the ‘walking on eggshells’ construction can be summed up as follows:

- phoric orientation (back and forward reference);
- conceptual orientation (marking the counter-thesis, marking common ground);
- balancing-out of emotional involvement: escalation (*doch/denn*) + mitigation (*eigentlich/auch*).

In the above, the construction shows itself to be a trace of social practice. This verbalization habit is completely inconspicuous in everyday communication but allows us to observe the way social experience solidifies into grammatical form in the seriality of language usage.

## 5. Doing it: digital discourse analysis

In the research sketch presented here I relied on information resources in different degrees of resolution and applied them to my research data: In linguistics, this involves validated knowledge of the language system, e.g. of the classification of parts of speech in German. Like many of my colleagues, I here trusted the classification system that is the basis of the Stuttgart-Tübingen tagset (STTS). Since I was interested in a part of speech which STTS does not classify separately, but which is established and recognized in the linguistic literature,<sup>36</sup> I performed my own classifications on the basis of the STTS-pre-structured data, fully relying on the institutionalized knowledge of my field. The same goes for a number of categories and terms taken for granted here, which I adopted without comment, such as ‘phrase’, ‘utterance’, ‘contribution’, or ‘coherence’. This is complemented by assumptions about

---

<sup>36</sup>Even though modal particles are sometimes termed *Abtönungspartikeln*, i.e. ‘mitigation particles’, primarily in the older tradition.

the informative value of the analyzed corpus. The task was to map the contemporary public debate on bioethics in Germany in a corpus in such a way that statements about it can be made. We were dealing with the question of representativeness here. We followed the standard assumption in the field, which says that the quantitative relation between sample and population studied under the term ‘representativeness’ in the empirical social sciences cannot be meaningfully addressed in corpus linguistics, because the population (here: the total quantity of all contemporary utterances in the German bioethics debate) is systematically unknown. Corpus linguistics therefore takes representativeness to mean the extent of correspondence between population and sample, based on the variation of the corpus data: “Representativeness refers to the extent to which a sample includes the full range of variability in a population.”<sup>37</sup> A study on the written language of contemporary German therefore should not be restricted to recipes and obituaries, but map out the entire spectrum of types of text. We put the corpus together on that principle.<sup>38</sup> What was also necessary for designing the corpus was a definition of ‘bioethics’ taken from philosophic literature,<sup>39</sup> as well as the question, to be asked in each individual case, whether an utterance falls into that category. As a decision-making tool we created a category grid of topics relevant to bioethics, based on specialist literature and our own pilot surveys. Then the attendant research group<sup>40</sup> spent one year recording relevant radio and TV programs and transcribing the directly uttered contributions to the debate they contained (quotes from reports and call-in formats, media discussions). On the basis of the topic list we also did a keyword-based collection of relevant threads in internet forums, position papers published online, as well as printed publications we considered relevant to bioethics. Wherever possible, we assigned role labels to speakers based on external labeling (e.g. descriptions of discussion programs in TV guides) or interactively validated self-positioning (e.g. positioning as a ‘layperson’ in online forums). Like all processes of corpus formation, this consisted of a perpetual succession of strictly method-driven procedures and workable decisions, which were made to the best of our knowledge and belief and discussed in research group meetings.

---

<sup>37</sup>Douglas Biber: *Representativeness in Corpus Design*. In *Literary and Linguistic Computing* 8(4) 1993, pp. 243–257.

<sup>38</sup>A comprehensive account and discussion of the creation of the Heidelberg bioethics corpus can be found in Müller, *Sprachliches Rollenverhalten* [same as footnote 4], pp. 125–149.

<sup>39</sup>“[...] the critical engagement with the moral dimensions in the contexts of action affected by the life sciences: biomedicine, biotechnologies, and ecology.” Quoted after Marcus Düwell, Christoph Hübenthal & Micha Werner (eds.): *Handbuch Ethik*. Stuttgart et. al.: J. B. Metzler 2002, p. 247.

<sup>40</sup>Apart from me, student assistants and colleagues involved in the construction and labeling of the Heidelberg bioethics corpus were: Thomas Bögel, Anja Chaluppa, Johanna Emmerich, Clara Herdeanu, Franziska Köder, Anna Mattfeldt, Jörn Stegmeier, and Friedemann Vogel, plus the colleagues from the media center of the library of Heidelberg University.

## 6. Only one step away: the text

I worked on the corpus with the standard methods of corpus linguistics.<sup>41</sup> Here, the most important instrument is also the simplest, namely the concordance. The possibility of treating the occurrence of a focus construction as a series whose syntagmatic environment can be described in variation and exemplariness is the basic resource of digital discourse analysis. The context-sensitive series is the determining factor that connects the individual record with statistic corpus results. The result that two-part modal particle constructions occur with significantly higher frequency in the bioethics corpus than in comparable corpora is neither a linguistic nor a discourse-analytical statement. The functionally dense description of a verbalization from an individual record in turn does not allow for statements on pattern formation and is thus again neither linguistically nor discourse-analytically sufficient. Another problematic procedure is testing hypotheses gained in the analysis of individual records quantitatively against the corpus, because it is not possible to show that the properties of the pattern, based on which we can reject a null hypothesis, are also the properties established via text analysis of the individual verbalization. In the end, only the concordance allows the control over the conditions of use of constructions within the discourse.

Corpus-linguistic research is largely based on the idea of the linguistic series in context, inasmuch as it understands language – with reference to Firth – as expression complexes situated in usage, serialized, and culturally contextualized.<sup>42</sup> With that said however, studies with an focus intrinsic in language structure can be distinguished from those that examine regularities in the phenomenal domains of ‘cognition’, ‘action’, ‘interaction’, ‘society’, or ‘empirical epistemology’ on the basis of authentic language use (spoken and written). This type of research was summarized under the label *corpus pragmatics*, by Felder, Müller, and Vogel:

We take corpus pragmatics to mean a linguistic research approach which examines the reciprocal relationship between linguistic means on the one hand and context factors on the other hand in digitally prepared corpora, and whose goal it is to establish a typology of form-function correlations. Such context factors can potentially affect the dimensions *action*, *society*, and *cognition*. Notably, the analysis makes use of a combination of qualitative and quantitative methods.<sup>43</sup>

---

<sup>41</sup>For an overview see e.g. Tony McEnery & Andrew Hardie: *Corpus Linguistics: Method, Theory and Practice*. (= Cambridge Textbooks in Linguistics) 1. publ., Cambridge [et.al.]: Cambridge Univ. Press 2012.

<sup>42</sup>John R. Firth: *Papers in Linguistics (1934–1951)*. Oxford: University Press 1957. Elena Tognini-Bonelli, Elena: *Corpus linguistics at work*. Amsterdam: John Benjamins 2001, p. 157 ff.

<sup>43</sup>Ekkehard Felder, Marcus Müller & Friedemann Vogel: Korpuspragmatik. Paradigma zwischen Handlung, Gesellschaft und Kognition. In Ekkehard Felder, Marcus Müller & Friedemann Vogel (eds.), *Korpuspragmatik. Thematische Corpora als Basis diskurslinguistischer Analysen*. Berlin, Boston: De Gruyter 2012, 3–30, p. 4.

Of course, corpus-pragmatic research is not limited to the interpretation of concordances. Apart from the co-occurrence analysis (including the calculation of co-occurrence networks<sup>44</sup>) and different forms of the analysis of specific, part specific, or lexically unspecific n-grams, the primarily relevant research in the thematically-oriented discourse analysis is that with keywords.<sup>45</sup> Those are words which occur in a data population A relative to a data population B with significantly higher frequency than would be expected if you assumed that all words are equally distributed in both populations. Moreover, this branch works with automatic, semiautomatic, and manual annotations of semantic and discourse-pragmatic categories (e.g. ‘defining’ or ‘arguing’). The French tradition of digital discourse analysis brought forth lexicometric methods.<sup>46</sup> In recent times, innovation is found primarily in the refining, recombination, variation, and visualization of these approaches<sup>47</sup> as well as in the integration of discourse-linguistic models, e.g. from frame semantics.<sup>48</sup> What these approaches have in common is that the digital accesses are understood as subprocesses of the analysis, whose results need to be contextualized driven by theory. This typically happens – explicitly or implicitly – via the main aspect of the discourse function of linguistic signs, a function which in turn can only be reconstructed on the individual record and its validity measured on the concordance. Therefore, meaningful digital discourse analysis is always only one step away from the text.

---

<sup>44</sup>Those are networks created by multiple iterations of co-occurrence analyses based on a focus construction. Cf. Friedemann Vogel: *Das LDA-Toolkit. Korpuslinguistisches Analyseinstrument für kontrastive Diskurs- und Imageanalysen in Forschung und Lehre*. In: *Zeitschrift für Angewandte Linguistik* 3, 2012, pp. 129–165, here p. 139, as well as Marcus Müller, *Sprachliches Rollenverhalten* [same as footnote 4], S. 168 ff.

<sup>45</sup>Cf. Paul Baker: *Using corpora in discourse analysis*. London, New York: continuum 2006.

<sup>46</sup>Cf. Ronny Scholz & Annika Mattissek: *Zwischen Exzellenz und Bildungsstreik. Lexikometrie als Methodik zur Ermittlung semantischer Makrostrukturen des Hochschulreformdiskurses*. In: Martin Nonhoff et. al. (eds.): *Diskursforschung. Ein interdisziplinäres Handbuch. Band II. Methoden und Analysepraxis. Perspektiven auf Hochschulreformdiskurse*. Bielefeld: transcript 2014, pp. 86–112.

<sup>47</sup>A good overview over current tendencies in discourse-analytically oriented corpus linguistics, especially in the realm of visualization, may be found on the website run by Noah Bubenhofer: <https://www.bubenhofer.com/> [last accessed on Nov. 13, 2017].

<sup>48</sup>Cf. Marcus Müller & Jörn Stegmeier: *Investigating risk, uncertainty and normativity* [same as footnote 3].