

MANFRED W. HELLMANN

## The Lexicographic Exploitation of the *Wendekorpus*: Looking back at the Themes and Discourses of the *Wende* with the *Wende- Wörterbuch*<sup>1</sup>

### 1 Foreword: Linguistics in Times of Revolution

Three times in the twentieth century Germans experienced the collapse of a German state: in 1918 the Kaiser's Empire, in 1945 the National Socialist Third Reich and in 1989-90 the German Democratic Republic. Each time the collapse was followed by a fundamental change of social system, the replacement of a large part of the ruling élite, a change in the dominant political and intellectual models and at least in the case of the fall of the GDR, or *Wende* – by a fundamental change in the economic system. In every case the upheaval extended beyond the socio-political and economic fields and deep into the everyday life of people, and influenced their linguistic usage.

Times of radical change like these are of particular interest to linguists: as a rule language change occurs more quickly than in more peaceful times. Past, present and future become entangled, political ideologies become entangled, political ideologies and their

1 This English language version (translated by Martin Wynne) is a reworked and slightly abridged version of the paper given in Canterbury in February 1999. A German language version has appeared as Hellmann, Manfred W.: 'Wörter in Texten der Wendezeit' 1989–1990 – Ein Wörterbuch zur lexikographischen Erschließung des 'Wendekorpus' in Jordanova, Ljubima (ed.): *10 godini promjana v Iztotschna Evropa (10 Jahre Wende in Osteuropa)*, (= Sociolinguistika Bd. 4), BULLEKS: Sofia (Bulgaria) 1999, S. 11–39.

corresponding linguistic usages fight for predominance. What had previously seemed uncontroversial suddenly comes under attack or becomes taboo – and *vice versa*. Groups whose voices were previously barely heard suddenly come to dominate the media, whilst others lose resonance with the public or fall silent. People's everyday lives change rapidly, along with their vocabulary and ways of expressing themselves. It is a legitimate interest – one might even call it a duty – of linguistics to monitor and describe this process.

This type of description of linguistic usage includes (but is not merely) the description of contemporary vocabulary, including both everyday words and the most salient keywords of the time. It can also make a contribution to the historical study of themes and discourses, and to the intellectual history of the time and of the speech community of the time, in this case the Germans.

Before the description of the language of the time can be attempted, linguists must make the language describable. For this, a principled selection of contemporary texts has to be made, and to make them usable by the anticipated research community, these texts have to be prepared as a corpus for linguistic investigation. The type and the quality of the possible scientific results depend to no uncertain extent on the type and quality of this text corpus.

It is regrettable that such corpora are not available for 1918–19 or for the post-war period 1945–49. It is understandable that they were not created at the time, but the fact that they have not subsequently been created at all is a serious defect. Today, to fail to document contemporary changes would be unforgivable.

In the autumn of 1989 the citizens of the GDR, with admirable courage, set about the task of first democratising and then removing their totalitarian system of government, dominated by the *Sozialistische Einheitspartei Deutschlands* (Socialist Unity party of Germany, or SED), a system officially known as *real existierender Sozialismus*. The will of the majority forced the unification of the GDR with the 'old' Federal Republic, and along the way fierce arguments raged and great difficulties were encountered, not only political and economic, which have still not been overcome. It is not surprising that when half of the people have never experienced any alternative to the system in which they were brought up, the linguistic habits of that

system are taken for granted. It is not so easy to reshape words as it is to reshape, for example, industry or the electoral process.

Times of revolutionary change have their own dynamic. They put people under pressure to act; not only the participants and those affected, but also those who have an interest as scientists in the events. Already during the time of the revolutionary changes in the GDR, which Germans along with Egon Krenz and the Politbüro of the SED learned to call the *Wende* (although this expression was already used differently in the public discourse of the FRG), there were plans to document and describe these changes, which also amounted to no less than a revolution in the public discourse of the GDR. It became the urgent duty of the Institut für Deutsche Sprache (IDS) in Mannheim, as the body responsible for the exploration of the contemporary German language in the FRG, to gather relevant texts as quickly as possible, and to document and make available electronically a linguistic corpus for its own and for external research.

## 2 The *Wendekorpus*

As early as 1991 the IDS (initially in tandem with the *Zentralinstitut für Sprachwissenschaft* at the GDR Academy of Sciences while it still existed, and then alone) had started the *Gesamtdeutsche Korpusinitiative* project. The end of 1992 saw production of the *Wendekorpus* (WK) as the result of this initiative. It is divided into two sections: (i) the *Wendekorpus DDR* (WKD) and (ii) the *Wendekorpus BRD* (WKB), with altogether 3.34 million running words of text. The WKB comprises approximately 55% of the total and the WKD approximately 45%. The following is an overview of the word counts:

Number of:	WKB	WKD	WK total
Texts	1,755	1,632	3,387
Sentences	105,779	99,647	205,426
Word types	106,621	79,374	141,233
Word tokens	1,794,000	1,546,000	3,340,000

(For further general information on the *Wendekorpus*, see Herberg and Stickel 1992 and for more detailed information on the construction of the corpus, and on its limitations, see Hellmann 1996, pp. 198–200).

In addition, some general information:

The *Wendekorpus* is not a representative synchronic corpus. Rather, it is a thematically-oriented corpus, in which the individual texts were selected according to two central themes:

- (i) From the protest movement to the democratisation of the GDR;
- (ii) The difficult road to German reunification.

The texts originate from the period May 1989 to the end of 1990, and within this time scale, the climax of the *Wende* from October 1989 to March 1990 is much more strongly represented than other phases. The strong points of the corpus are the representation of certain events, such as the celebrations of the 40th anniversary of the GDR (7 October 1989), the demonstrations and the reaction of the state forces, resignations of prominent people, the liquidation of the *Stasi*, the opening of the Wall (9 November 1989), the parliamentary (*Volkskammer*) vote of March 1990; and later the economic and currency union (July 1990), the unification negotiations in the run-up to reunification on 3 October 1990. All of the texts, despite their thematic diversity, fit into the aforementioned two central themes, and essentially into only these two.

As the period covered by the corpus ends in November to December 1990, the developments of 1991 and later are not properly covered.<sup>2</sup>

There are a wide variety of text types in the corpus: alongside reports and commentaries from the press, and parliamentary transcripts from both the *Volkskammer* and the *Bundestag*, there are official proclamations, reactions from the public (e.g. correspondence), leaflets

2 For this reason the words *evaluieren* / *Evaluierung* / *Evaluation* and *abwickeln* / *Abwicklung*, which later took on an important negative connotation, have few examples in the corpus.

and election pamphlets, speeches and interviews, broadcast appeals, manifestos and more.

On the other hand, it lacks texts from the spoken language, domains such as sport or international politics, and the non-editorial sections of newspapers (e.g. job and flat adverts, personal, commercial advertising etc.).

It was not the job of the researchers in Mannheim and East Berlin who selected the texts to construct a corpus that was somehow representative of a particular area, or balanced with respect to its constituent parts or text types. Much more important was the goal of documenting the themes and discourses of the *Wende* through the variety of texts with all their emotions, contradictions, repetitions and interconnections – in other words, to realise a discourse-oriented approach.

The corpus builders certainly attained their goal. It is possible to criticise the corpus and the criteria for its construction on several points, for it lacks some features which a corpus user would like to have available. It is also questionable whether it is sound from a methodological point of view to construct a corpus for the purpose of addressing such a particular research issue. But if the original premise of the corpus builders is accepted, it follows that the corpus is an outstandingly rich resource for the themes and discourses of the *Wende*, and just as rich in the density of relevant vocabulary as in the quantity and quality of the examples of usage of the individual words.

### 3 The Format of the Corpus Texts

The corpus texts are available in electronic form, encoded according to the 'Mannheim Text Conventions', which enable them to be used with the COSMAS system. The following conventions are followed:

- all punctuation characters are separated from the word by a space character, and every sentence is terminated with a full stop (between spaces);

- headings and other emphasised text elements are marked;
- discontinuous hyphenated words are concatenated as *Unterstrichkomposita* ('underscore compounds'): *Partei- und Staatsführung* appears in the examples as *Partei\_führung* und *Staatsführung*.

As these peculiarities would make the examples difficult to read when they are presented in a dictionary, they have to be edited out by hand or with simple automatic routines. In this way the final text is restored to something like 90% of its appearance before encoding.

#### 4 COSMAS<sup>3</sup>

In parallel with the collection and preparation of the *Wendekorpus* the COSMAS (CORpus Storage and MAintenance System) was developed. This system allows the user to search the *Wendekorpus* (and all other IDS corpora) for words and phrases, not only at the level of word-forms, but also lemmas, and to output the resulting examples in different formats and with variable context. Tables of words and frequencies can also be generated.

This useful search system has its strong and weak points, which I have already reported on elsewhere (Hellmann 1996, pp. 202–204). Here the following points are relevant:

In the output from a search query COSMAS gives information on the name and date (if known) of the source text for every output line. In this way, nearly all the examples can be dated and their source identified (and usually also therefore the speech community responsible for producing the text). In addition, the texts are organised in six time periods, or phases, which relate to the main events of the *Wende*

3 COSMAS and many text corpora held by the IDS are available for external users' on the internet at <http://corpora.ids-mannheim.de/cosmas/>. Further information can be found in the handbook developed in-house at the IDS (al-Wadi 1994).

(cf. the overview by Herberg, Steffens and Tellenbach, 1997, p. 2). The COSMAS software is sensitive to the time phase codes, and the search results can therefore be sorted not only alphabetically but also according to these phases. This is important when users want to search for changes in usage taking place during the period of the *Wende*.

*Phase index and word frequencies:*

On the basis of this time phase tagging of all the texts in the corpus, a 'phase index' (*Phasenregister*) was made, in which every word in the corpus is arranged with its frequency in each time phase, along with its total frequency in the WKI, total frequency in the WKB and the total frequency in the WK. See section 6.8 below for more information on the word frequency tables.

## 5 Three projects to analyse the *Wendekorpus*

Within the framework of the *Sprachwandel der Wendezeit* project, three subsidiary projects were started in the IDS to analyse the corpus:

- 1 *Lexicological-lexicographical analysis of vocabulary changes specific to the Wende ('Keywords')*:  
The resulting publication was Herberg, Steffens and Tellenbach (1997).
- 2 *Meaning variation in texts on German Unity*  
Resulting publications were Fraas (1996) and Steyer (1996).
- 3 *Documentary and lexicographical investigation of the Wende corpus ('the Wende-dictionary')*  
Resulting publications were Hellmann, Melk and Nikitopoulos (2000) and a bibliography to parallel the lexicographic work: Hellmann (1999).

These three projects clearly differ in their aims, methodology and results. This is intentional. It is the precise aim of the heterogeneity of the approach to guarantee the visibility of the language of the *Wende* in all its variety and complexity. The unifying factor is that the language of the *Wende* (and the *Wendekorpus*) cannot adequately be described with single-word-based lexicography alone.

However, the traditional lexicographical approach is not dispensable. An alphabetically sorted list of words is still the most practical method of accessing the riches of the corpus and analysing the *Wende*. In this respect the third project, the exploitation of the corpus for the *Wende-Wörterbuch*, will complement the other two, and is not in competition with them.

None of the three projects claims to analyse the corpus exhaustively. In all aspects the corpus contains more than has been worked on in the IDS, and much more than can be presented here. There is plenty of scope for further research.

## 6 The *Wende-Wörterbuch* project

### 6.1 Aims

The aim of the project, according to the IDS workplan, was the drawing up of an alphabetical, corpus-based word-list of around 1500 headwords and sub-entries with commentaries on their relation to the themes and discourses of the *Wende* as they occur in the corpus. For each keyword, concise examples were given of typical usages from the different sources and time phases. In addition, information on frequency of occurrence (East and West) for all inflected forms, compounds and derived forms of the keyword was included. Explanatory notes were limited to cases where the examples could not adequately show the specific meaning.

Cross-references make links between thematically close words and add an additional level of analysis of the themes and discourses of the texts.

The dictionary is aimed primarily at users interested in the vocabulary and usage of the *Wende*, and particularly at those who do not have direct access to the *Wendekorpus* on computer, and also at those who want fast word-based access to the themes and discourses of the period.

The work began with the pilot project in 1993. At first we expected to find mainly examples of semantic change, neologisms, new meanings and new usages. These expectations proved to be only partially correct. Lexical neologisms, and changes in word meaning, which can be satisfactorily described at the word level with the help of the alphabetically organised dictionary, occurred far less frequently than expected during the *Wende*. Here are some examples of neologisms:

- *Mauerspecht, Trabiklatschen, Botschaftsflüchtling, Vereinigungskriminalität, Treuhandanstalt, Beschäftigungsgesellschaft;*

Examples of new meanings of words (*Neosemanteme*):

- *Runder Tisch* ['round table'] meaning '*basisdemokratisch bestimmtes Mitregierungsgremium auch oppositioneller Gruppen*', ['talks on basic democratic reforms involving government and opposition groups'];
- *Abwicklung* ['liquidation (of a company)'] meaning '*Auflösung auch wissenschaftlicher, kultureller, sozialer Einrichtungen der DDR*' ['dissolving of scientific, cultural and social institutions of the GDR'], but also in referring to people, meaning '*aus dem Arbeitsverhältnis entfernen*' ['sacking'];
- *Wendehals* ['wryneck (a bird that can turn its head 180°), or a turncoat'] meaning '*Mensch, der seine politische Meinung opportunistisch (während der Wende) blitzschnell ändert*' ['person who quickly and opportunistically changes political allegiance during the *Wende*'];
- *Warteschleife* ['turning loop, or holding pattern (of an aircraft)'] meaning '*(bezahlter) Wartezustand zwischen Entlassung aus dem Staatsdienst der DDR und der Wiedereinstellung bzw. der*

*Arbeitslosigkeit*, ['Fully-paid time between jobs after redundancy from GDR state employment and before re-employment or unemployment'];

- *Blockflöte* ['a recorder (musical instrument)<sup>4</sup>] ironic or mocking, meaning '*Angehöriger der von der SED beherrschten (ehemaligen) Blockparteien der DDR*' ['member of one of the non-communist parties controlled by the SED'] etc.

These neologisms coined during the *Wendezeit* number in total approximately one hundred. There are certainly interesting, but they do not by any means exhaust the possibilities of exploring the *Wendekorpus*.

What predominates in the corpus in a quantitative sense is not such words but rather words like *Verantwortung* and *Würde*, *Nation* and *Einheit*, *Souveränität* and *Vereinigung*, *Dialog* and *Wende*, *Erneuerung* and *Umgestaltung*, *Partei* and *Sozialismus*, *Demokratie* and *stalinistisch*, *bankrott* and *real existierend*, *Täter* and *Opfer*, *Stasi* and *Akte*, *Aufbruch* and *Zusammenbruch*, *Menschen* and *Volk*.

What, for example, changed in the meaning of *Volk*? Up until the *Wende* the SED proclaimed the '*Einheit von Partei und Volk (der DDR)*' under their leadership, the '*Führung der Partei der Arbeiterklasse*'. With the cry '*Wir sind das Volk!*' (with the emphasis on '*Wir*') the demonstrators did not create a new meaning. Rather they challenged the Party which claimed unity and claimed an identity of wishes and interests between them and the citizens. Not '*ihr seid*' but '*wir sind das Volk!*' the people, from whom, in a democracy, all state power should flow.

When, at the beginning of December 1989 in place of this slogan another appeared – '*Wir sind ein Volk!*' (with the emphasis on '*ein*') – the demonstrators made current a different meaning of *Volk* from that above, namely something like '*staatsübergreifende Gemeinschaft von Bürgern auf der Grundlage gemeinsamer Nationalität*' ('a community of people across state boundaries and united by common nationality'). But that is really not a new meaning, but one which was

4 Flöte also already had negative meanings when applied to people, which could be translated as 'wet blanket', 'drip' or 'jerk'.

always active in public discourse in the FRG, and also indirectly therefore in the GDR because of its negation.

The difference is to be found in the role that this word and its meanings came to play in the changing discourse of the autumn of 1989. For the SED the first slogan cited above is an alarming 'counter-revolutionary' act 'from below' which challenges their legitimacy as the leading power in society. The second slogan calls into question the existence of the GDR as a state by turning upside down the concept of the *sozialistische Nation (deutscher Nationalität)*, and thereby calls into question their right to rule.

These changes have little to do with lexical semantics, but relate more to the revolutionary emergence and flourishing of new previously suppressed or regulated themes of discourse – and as such they are exciting objects of analysis. The same could be said for most of the other words cited above, for example *Dialog*.

It was above all the church and associated groups who challenged the SED in September and the beginning of October 1989, and in the face of their resistance (and *Sprachlosigkeit*) offered *offenen Dialog* even with *Andersdenkende*. Under pressure from growing demonstrations the SED then called on October 18 for *breiten gesellschaftlichen Dialog mit allen (positiven) Kräften*. Perhaps this was the '*letzten sprachlichen Selbstrettungsversuch* [last attempt to save themselves by talking]' (Good 1991).

Citizens' groups then fought from the end of October against the dialogue becoming an end in itself, and against delay in the processes of *revolutionäre Erneuerung, durchgreifende / grundlegende Veränderungen, der Aufbruch / der wirkliche Durchbruch zur Demokratie*. Throughout, the meaning of *Dialog* was not the issue, but much more importantly it was the control of the themes of discourse, or to be more precise, about who was determining the dominant themes and the key words of the discourse, and whether it was to be the citizens movements and the demonstrators.

A dictionary which wishes to fully exploit a corpus must first take into consideration what is specific and special to this corpus and the texts it contains, and develop a method of explaining the words which will make it possible for the reader to gain access via the words to precisely these specific facts.

In any case a dictionary of this type can, when it is based on a corpus, only exploit what is there in the corpus.

## 6.2 *The lists of headwords*

In the beginning of the processing there was only an initial list of around one hundred words, which could be safely assumed to have played an important role during the *Wende*. From this initial list some 200 search strings were chosen, which were then used to launch the first COSMAS search queries.

The results of these searches led to numerous new candidates for headword status, arising sometimes from paradigmatic relations (compounds, derived forms, synonyms and antonyms), and sometimes from syntagmatic relations (syntactic connections, collocations). Furthermore, sometimes new candidates came from contextual relations; every batch of new search results brought new pointers to related themes and with them new vocabulary, or with the very same themes, words and phrases characteristic of a specific group of people. Often these different aspects were blended together, as the example of the word *Wende* shows below.

As near-synonyms of different types the following were used for *Wende*: *Umgestaltung, Perestroika, Umbau, Umbruch, Umschwung, Umsturz, (revolutionäre) Erneuerung, (grundlegend/durchgreifende) Wende/Veränderung (en), Aufbruch, (friedliche/sanfte) Revolution, Zusammenbruch (des alten Systems/der SEDHerrschaft)* and more.

For *bankrott* there were adjectives and adverbs used with negative connotations to label the GDR and its economic system: *verrottet, verschlissen, verkrustet, desolat, heruntergewirtschaftet, abgewirtschaftet, konkursreif, marode, diskreditiert, miserabel, kaputt*.

For *Dialog*, an analysis of several GDR newspapers from the end of 1989 up until the highpoint of the SED's *Dialog-Offensive*, shows 32 different near-synonymous expressions, both nominal and verbal, and often with a distinctively very contemporary resonance (see Hellmann 1993).

A year later the candidate list of headwords had reached nearly 2000 entries, and after another year around 4000 - a number which clearly exceeded the capabilities of the research team to process. So a line had to be drawn here.

At first a lower threshold was introduced. Words which occurred less than 5 times were excluded, with a few exceptions. For main headwords the threshold was set at 10 occurrences. It was decided to include all words which would assist in understanding the themes and discourses of the *Wende*. The crucial factor for word selection was not therefore the semantic properties of the words (e.g. changes in meaning), nor whether their use was restricted to East or West (e.g. although in the Eastern texts *Zielstellung* was used more frequently, while *Zielsetzung* was preferred in the West, this fact was not relevant). Not even a new word coined in the *Wende* could be guaranteed a place in the list. The basic criterion for inclusion was the significance of a word in opening up to the reader the themes and discourses of the *Wende*.

Multi-word phrases were also included as headwords, for example *chinesische Lösung*, *sozial abfedern*, *Grundordnung (freiheitlichdemokratische)* and whole clauses like *Wir sind das Volk, zu spät kommen – bestrafen*, also word components like *Ex-*, *-schaffend*, *Zentral-*, which appear in the dictionary to show how productive they were during the *Wende* (see also section 6.7 below).

Further restrictions were required. After it became clearer in the second half of the project which Schlüsselwörter were being used in the sister project (see section 5 above), these areas of vocabulary were given a lower priority.<sup>5</sup> Instead, other fields were considered, for example the vocabulary of emotion, which was astonishingly<sup>6</sup> extensively used in the corpus.

After this narrowing down, there still remained some 2100 headwords which had earned an entry in the dictionary, according to the criteria outlined here. Of these, around 1980 were given a more or less

5 This was due not only to the necessary restriction in the number of headwords, but rather more to the fact that an orientation based on vocabulary fields is able to represent connections in a much clearer form than in an alphabetical dictionary which, by necessity, isolates words.

6 This was based on a separate examination of almost 900 words dealing with emotion and morality in the *Wendekorpus* (see Hellmann 1997b). This area of vocabulary is richly represented although it was not a determining criterion the selection of texts for the corpus. Around 30% of these words are included in the dictionary.

intensive lexicographic treatment. The rest fell victim to the time constraints of finishing the dictionary.

### 6.3 Problems of lexical density and the wealth of examples

It was not only the density of *Wenderelevant* vocabulary in the corpus that caused difficulties, but at least as problematic was the overwhelming number of examples resulting from the COSMAS searches for the selected headwords. For many words this number unfortunately far exceeded expectations.

The following is a selection of some large files generated by the searches, in ascending order of frequency (in brackets the numbers of different types and tokens):

*Verantwortung* (not *verantwortungslos*) (47 / 1467),

*Einheit* (not *einheitlich*) (191 / 2529),

*Nation/national* (252 / 3165),

*Mensch* (not *menschlich*) (205 / 4844),

*Bürger* (not *bürgerlich*) (290 / 6353),

*sozial* (not *sozialistisch*) (499 / 6884),

*Volk* (not *völkisch, bevölkern*), (501 / 7124),

*Partei* (not *parteilich*) (839 / 7477),

*Staat* (not *staatlich*) (954 / 8895),

*Deutsch/Deutschland* (412 / 17499),

*DDR* (with all hyphenated forms) (1941 / 23889).

The last file was unworkable in the form given above, not least owing to computer storage limitations. Instead, only searches based on certain selected forms were run and processed, such as *-DDR* (e.g. *Ex-DDR, Noch-DDR, BRDDR*) and *DDR-Bürger*. With *Staat*, words which existed on their own as keywords were excluded (e.g. *Staatsbürger, Staatssicherheit, Staatsvertrag*), others were eliminated as irrelevant (e.g. *Staatsanwalt, Staatssekretär*), and the rest only partially checked, for example, blocks of 10 to 20 were selected every 100 or 200 examples. Even this is very time-consuming, as the simple calculation below shows.

Starting with a COSMAS search output of 1000 hits, if the lexicographer takes only 8 seconds to read an example on the screen

(this is on average 10 to 15 lines of text), and another 8 seconds if he needs to mark it as interesting or make some other note, then a set of 'only' 1000 examples requires an average duration of four and a half hours. For entries with more than 200 hits, two readings are necessary, and for those with more than a thousand three passes are needed, in order to group the examples and to select one of the examples as suitable for citation. This requires an effort of two working days or more. And this represents only the preparatory stage of the writing of the dictionary entry, with the insertion of the selected examples and all the other editorial and technical stages yet to be done.

#### *6.4 Methodology: from headword to word entry*

Given these circumstances and the severely limited personnel (only two researchers), a methodology had to be developed to cope with both the large number of headwords and the overwhelming abundance of examples per headword.

The following procedures were decided on:

- the simplest possible structure of the word entry; summarisation of several (derived) headwords under one lemma; categorisation of compounds as subentries under a headword;
- complete renunciation of definitions and other information normally given in a dictionary, e.g. grammatical information;
- instead, concise, summarising commentaries on usage, and if necessary examples are cited, focussing on the themes and discourses to which the word was connected, by whom it was used and how (as far as could be ascertained);

in addition, supporting evidence of the usage is given in word frequency tables (including compounds and derived forms).

#### *6.5 The structure of a dictionary entry*

A dictionary entry has, in principle, a very simple structure. It has five parts, or fields:

- 1 Lemma (bold and underlined) with optional additional information like 'Abk. für...';
- 2 optional information on words forms or meanings not dealt with in the dictionary entry, for example, under *Angst* in brackets '(ängstigen, ängstlich)'. These words were not searched for in COSMAS and not dealt with in the word entry;
- 3 Editorial commentary (field name: *Komm*). The field *Komm* contains the real description of the use of the headword(s) in the *Wendekorpus* texts. There are commentaries of 4 different types, reflecting different levels of analysis:
  - short prose commentary on the headword;
  - prose commentary summarising all of the usages which appeared in the examples;
  - prose commentary with individual example concordances for clarification of different usages;
  - commentary almost completely comprising citations, with numerous examples, and a relatively small amount of prose commentary.
- 4 Additional cross-references (field name: *Verw*).
- 5 A summary line with basic information on frequency of occurrence, with a breakdown of the respective frequencies for the whole corpus (*WK*), the Western section (*WKB*) and the Eastern section (*WKD*).

These five fields together make up the word entry.

In addition there is a word frequency table for each main dictionary entry. This table contains a list of all word-forms (inflected forms; compounds, derived forms) found by COSMAS along with the frequencies in *WK*, *WKB* und *WM* (see also section 6.8 below.)

The word frequency tables are detached from the word entries and printed (in two columns) as an appendix to the dictionary.

### 6.6 Stages in the compilation of an entry

The process of creating an entry has several stages, some recursive, which are described below in a simplified form:

- 1 The selection of the lemma (headword or headwords) and the selection of the (most efficient) COSMAS search string;
- 2 COSMAS search, creation of the file of examples and checking; revision of the examples (i.e. the reconstruction of the 'normal' writing conventions);
- 3 Initial analysis of the examples, creation of a first draft for the structure of the entry (main headword, sub-entries);
- 4 Second analysis of the examples, refinement of the draft for the structure of the entry; selection of examples to be used in the dictionary;
- 5 Creation of the framework of the entry, with main sections; arrangement of the primary and secondary structure;
- 6 Filling in the slots in the structural framework: addition of quoted examples in the framework; re-working of the selected examples (abridgement, syntactic adjustments); completion of the entry with structural or contextual commentary by the lexicographer;
- 7 Filling in of cross-references (→) in the entry and/or in the reference field (*Verw*);
- 8 Revision of the entries;
- 9 Insertion of the finished entry into the alphabetical ordering of the dictionary;
- 10 Incorporation of the summary line, including the basic frequency information (taken from the word frequency table).

The main lexicographic work focuses naturally on steps 4 to 6.

The commentary in an entry aims to answer the following central questions: *Who uses the word in what way, during which time period, in what sorts of texts, and in relation to which themes and discourses?*

These central questions mean that there are several different aspects which have to be taken into account in the structure of the commentary. Some of them are based on categories which are external to the text, others are text-internal and yet others based on semantic categorisation.

The most important are the following three aspects:

- 1 The semantic structure of the headword is determined by attempting to answer the following questions:

- Does it have only one meaning, or several 'main senses' which can be clearly distinguished? Is this semantic structure reflected in the corpus texts and the themes and discourses there?
- 2 Aspects external to the text can be summarised as relating to three categories: the time phase, the speech community and the text type:
- (a) *Do the examples indicate a usage or change of usage which is specific to the phases and events of the Wende? Do certain events define the usage?*
- (b) *Is the use predominantly specific to the respective speech communities, in the East and West? (e.g. SED, or citizens' groups in the GDR, or the Volkskammer, or Western politicians in the Bundestag, Chancellor Kohl, etc.)*
- (c) *Is the usage characteristic of particular text types? (e.g. slogans on demonstrations, or transcripts of the Volkskammer, or election campaign pamphlets).*
- 3 Aspects internal to the text:
- These are above all related to the topic of the text: *In relation to which themes and in which speech situations is the headword primarily used? Is any change discernible? To what extent are the topics influenced by the time period, the speech communities or text types?*

Generally several of these aspects operate simultaneously to create the specific form of the texts, and thus the examples in the dictionary. The problem is that the compiler must decide to which criterion he will assign primary, secondary or tertiary importance in the structure of the entry. If a headword has, for instance, several clearly distinguishable main senses, then the compiler can make these meanings a criterion for the divisions of the main structure. In so doing, however, other thematic aspects, some of which bear central importance, may unfortunately be assigned too little significance. The decision must be taken in each individual instance: often there is no perfect solution. Overlaps are also unavoidable.

The structure used most often consists of a combination of ordering according to rough temporal sequence and/or speech communities.

The criteria are marked by brief metacommentary in square brackets, usually at the start of a subdivision: e.g. [*vor der Wende, SED:*], [*während der Wende, Bürgergruppen:*], [*nach der Maueröffnung:*], [*vor der März-Wahl:*], [*nach der Währungsunion, West-Texte:*] and so on; further ordering by theme is then possible within this structure.

It is also common, however, that the primary structure is organised according to topic, and then the other criteria are secondary categories. Information about the topic is located at the start of a subsection, followed by a colon, and without further marking. The temporal aspect is always present in that all quoted examples include details of month and year of occurrence, and also the day in special cases.

There are, in addition, contextual comments designed to facilitate fuller understanding of the examples, that is, their co-text (which has of necessity been abridged).

When the main structure is decided, explanatory information is added to the individual sections. This can, as explained above, be a prose commentary or a set of corpus examples, or a mixture of the two. The most detailed and for the reader probably most informative form is the corpus examples, but this is also the most labour-intensive and lengthy.

The examples which were marked in the search output are then looked up. A suitable subset is selected and slotted into the entry, along with date and source information. As a rule some kind of revision is necessary, usually an abridgement, and often also a syntactic reordering or the addition of further remarks.

The following gives the example of the stages of the revision of the citations in the entry for *Lohn*.

Unabridged version (with 3 sentences of context):

W2B/BT1.50014, Bundestagsprotokolle (2. 1Jj. 1989); Sitzung Nr. 173, Bd. 15 1, S. 13010-13059, 89.11.08, S. 13012

Sie wollen nicht länger von politischer Mitbestimmung und Mitverantwortung ausgeschlossen sein. Sie wollen nicht unter persönlichen und wirtschaftlichen Bedingungen leben müssen, die ein von ihnen nicht gewolltes politisches System ihnen auferlegt, ein System, das ihnen sowohl persönliche Freiheit als

auch einen gerechten Lohn ihrer täglichen Arbeit vorenthält. Unsere Landsleute in der DDR wollen endlich selbst frei entscheiden können.

### Abridged and syntactically altered version:

[West-Texte, vor der Wende:] die Bürger in der DDR «wollen nicht ... unter einem von ihnen nicht gewollten politischen System ... leben müssen, das ihnen sowohl persönliche → Freiheit als auch einen → gerechten L. ihrer täglichen → Arbeit vorenthält» [Bu-Prot. 11/891];

### Further abridgement is also possible:

DDR-Bürger «wollen nicht unter einem. ... politischen System leben müssen, das ihnen ... einen → gerechten L. ihrer täglichen → Arbeit vorenthält» [Bu-Prot. 11/891];

### As it appears within the prose commentary:

[Vorwurf aus West-Sicht:] DDR-System enthält seinen Bürgern → gerechten L. ihrer täglichen → Arbeit vor

### In the headword commentary:

[kritisch zum DDR-System:] kein gerechter L. für geleistete → Arbeit; oder: Kritik an ungerechtem L. im DDR-System;

As can be seen, brief metacomments are placed before the quoted examples; the quoted examples have opening and closing angled brackets (« ... »); large omissions are marked with '...', the additional remarks of the compiler for enhanced understanding are in brackets (=...). The codes used to indicate the source (here: *Bu-Prot.* for *Bundestags-Protokolle*), including month and year, are in square brackets at the end of the entry. There is no source information given with prose commentaries.

If words occur in the examples quoted or in the explanatory information which are themselves headwords, they are marked with an arrow. In the example above this is the case with *Freiheit*, *gerecht* and *Arbeit*. All three (and many others in the entry *Lohn*) are headwords.

Such cross-reference words are only marked on their first occurrence in an entry.

If it is necessary to make reference to other headwords which do not occur in the entry, but are nevertheless interesting to compare, then these are entered in the cross-references field *Verweise* (*Verw.*):

Together, these two methods of cross-referencing show the strong links between large groups of headwords, and at the same time they show which discourse themes are especially well covered.

## 6.7 Types of dictionary entry

The number of different types of dictionary entry was kept as low as possible. There are the following types:

### 1 Overview headwords

With minimal commentary a mere overview is given of the productivity, as demonstrated in the word frequency tables, of the lemma in all its forms: inflections, compounds, derivations and orthographic variants. This type of entry is used particularly for word components like *Ex-*, *-schaffend*, *Zentral-* and *zusammen-*, but also with headwords like *Parteien-*, in which case they are usually followed by several sub-entries with normal commentary. Overview headwords have the same status as main headwords.

### 2 Main headwords

The lemma of the main headword is marked by bold type and underlining. The size of the entry and the amount of internal subdivision vary greatly. Where there is a rich internal structure, the divisions may be marked with numbering or the character '♦'. Main headwords always have a word frequency table.

Several headwords can be categorised as sub-entries under the lemma of a main headword entry. The relationship between these words and the main headword is shown typographically, as is explained below.

Derived forms are separated by forward slash characters (/). The base word form (as used for sorting the entries) is not usually repeated,

but replaced by a dash. In the case of a functional conversion (e.g. verb to noun, or noun to adjective), this is marked by a full stop before the dash. An example is the dictionary entry *Bürokrat, Bürokratie, bürokratisch*. This is written as Bürokrat/ -kratie/ .-kratisch. The words *Bürokratismus, Entbürokratisierung* and *unbürokratisch* are appended as subentries. The example of dableiben/ Dableiber/ dageblieben demonstrates another way in which this can be written.

Variant compounds based on the headword are separated by a semicolon, e.g. *Elend; Verelendung, Zehn-Punkte (-Plan, -Programm)*.

Attributive constructions are given either in their original order (e.g. *himmlischer Frieden (Platz des -)*), or with a different alphabetic ordering and separated with a comma (e.g. *Frage, deutsche.*)

Phrases are generally selected verbatim (e.g. *Wir sind das Volk, Einheit der Nation*), although they are occasionally shortened (e.g. *zu spät kommen – bestrafen* from 'Wer zu spät kommt, den bestraft das Leben' or *hüben – drüben* for 'hüben und drüben', 'von hüben nach drüben').

### 3 Sub-entries

These are mostly compounds or forms derived from the headword. They are marked in bold without underlining. They are separated on the page from the headword commentary by blank lines. Their internal structure is limited. They have no word frequency table, because the occurrences are accounted for in the table for the headword.

### 4 Cross-references between headwords

Cross-references are given in the following cases:

- between main headwords, especially with multi-word units: for the catchphrase *Kontinuität und Erneuerung*, a cross-reference is given under the entry for *Erneuerung* as:

*Erneuerung, Kontinuität und* → *Kontinuität und Erneuerung*  
under *Volk* cross-reference is given to the main entry *Wir sind das Volk*;

- with sub-entries which are listed outside of the normal alphabetic order: the keyword *unbürokratisch* is entered as a sub-entry under *Bürokrat/ -kratie/ -kratisch*, and in the normal alphabetic order under **U** there is the cross-reference:

*unbürokratisch – Bürokrat/ -kratie/ -kratisch*

*Keine Gewalt* is entered as a subentry under *Gewalt*, and so there is a corresponding reference under *K*:

*Keine Gewalt* → *Gewalt*

## 6.8 The word frequency tables

As a frequency list of all word-forms in the *Wendekorpus* will not be generally available, the word frequency tables of this dictionary are the only opportunity to find out about frequencies, although of course only for headwords in the dictionary.

The word frequency tables give complete information about morphological variation, productivity in the formation of derivations and compounds and, most importantly of all, the frequency of occurrence of all forms of the word (and of the lemma), broken down into overall frequency in the corpus, frequency in the Western texts and frequency in the Eastern texts.

The frequency information for West and East texts does not come from the COSMAS searches. It is extracted with a special program using the time phase index (see section 4 above) and the frequencies are added to the word forms in the table. These frequency counts are then added up in the final editing stage and given as the summary line at the end of the entry.

For multi-word phrase units like *aufrecht gehen* and *Keine Gewalt*, the total number of occurrences of the phrase is recorded (under 'Summe:'). There is no East/West breakdown though, as COSMAS cannot provide it.

Words occurring in a frequency table which are categorised as *wenderelevant* according to frequency (at least 5 occurrences) and verification of the context are marked with an asterisk. They may be, but are not necessarily, processed as headwords or sub-entries. In each case it is the base form which is marked (or the inflected form morphologically closest to the base form).

In order to make the longer tables easier to read, the inflected forms are lemmatised and the frequency counts added to those for the base form. The inflected forms remain visible. As there was not an

automatic routine available for lemmatising or adding up the frequencies, this had to be carried out manually.

## 7 Results

The dictionary is perhaps unusual in that it explores topics and discourses from a fascinating period, according to the evidence found in the *Wendekorpus*, rather than primarily recording word meaning and semantic change. Readers may judge for themselves whether and to what extent usage is described and explained.

Many comprehensive reference works and dictionaries of both meaning and usage already describe the period of the division of Germany, both political and linguistic, and reference is made to them here (see Select Bibliography below). The dictionary described here does not attempt to compete with such reference works, much less with intensive studies of specially selected vocabulary fields of the *Wendezeit*. Rather, it has its own purpose and place.

Over and above its practical uses, the headwords in the dictionary can be used, if so desired, as a key to unlock memories of a time which would be too soon forgotten if not captured in textual form in the *Wendekorpus*.

This work on texts from the *Wendezeit* was for us, although it lies only fifteen years in the past, also a journey of re-discovery and refreshing of memories. If the readers of the dictionary find themselves, in a similar way, reliving this time of radical change, of longed-for liberation, of heated argument, great hope and some disappointment, then an important, albeit unofficial, aim will have been fulfilled.

## Bibliography

- al-Wadi, D. (1994) *COSMAS Benutzerhandbuch* (Handbook for version R. 1. 3-1.). Mannheim, Institut für deutsche Sprache.
- Barz, I. & Fix, U. unter Mitarbeit von Schröder, M. (Eds.) (1997) *Deutsch-deutsche Kommunikationserfahrungen im arbeitsweltlichen Alltag*. (= Sprache – Literatur und Geschichte. Studien zur Linguistik/Germanistik Bd. 16), Heidelberg, Universitätsverl. C. Winter.
- Clyne, Michael G. (1995) *German in a changing Europe*. Cambridge, Cambridge University Press. [esp. Ch. 3: German in divided and unified Germany, S. 66–88.]
- Fraas, C. & Steyer, K. (1992) 'Sprache der Wende - Wende der Sprache? Beharrungsvermögen und Dynamik von Strukturen im öffentlichen Sprachgebrauch.' *Deutsche Sprache* Jg. 20, H. 2, pp. 172-184.
- Fraas, C. (1996) *Gebrauchswandel und Bedeutungsvarianz in Textnetzen – Die Konzepte IDENTITÄT und DEUTSCHE im Diskurs zur deutschen Einheit*. (= Studien zur deutschen Sprache Bd. 3). Tübingen, Narr.
- Good, C. (1991) 'Der Kampf geht weiter oder Die sprachlichen Selbstrettungsversuche des SED-Staates.' *Sprache und Literatur in Wissenschaft und Unterricht*, 67, pp. 48–55.
- Hellmann, M.W. (1990) 'DDR-Sprachgebrauch nach der Wende – eine erste Bestandsaufnahme.' *Muttersprache* Jg. 100, H. 2/3, pp. 266–286.
- Hellmann, M.W. (1993) 'Die Leipziger Volkszeitung vom 27. 10. 1989 – eine Zeitung im Umbruch.' In: *Muttersprache* Bd. 103, H. 3, [Themenheft *Sprache nach der Wende*] pp. 186–218 (Appendix pp. 214–218).
- Hellmann, M.W. (1994) 'Ostdeutsch-Westdeutsch im Kontakt – Brücke oder Schranke der Verständigung?' *Terminologie et Traduction* (Kommission der Europäischen Gemeinschaft Übersetzungsdienst – Luxemburg) No. 1, pp. 105–138.
- Hellmann, M.W. (1996) 'Lexikographische Erschließung des Wendekorpus – Ein Werkstattbericht.' In: Weber, Nico (Hg.): *Semantik, Lexikographie und Computeranwendungen* (= Sprache und Information). Tübingen, Niemeyer, pp. 195-216.
- Hellmann, M.W. (1997a) 'Tendenzen der sprachlichen Entwicklung seit 1989 im Spiegel der Forschung.' *Der Deutschunterricht* H. 1 [Themenheft *Sprachwandel nach 1989*], S. 1732.
- Hellmann, M.W. (1997b) 'Wörter der Emotionalität und Moralität in Texten der Wendezeit – Sprachliche Revolution oder Kommunikationsbarriere?'

- In: Barz, I. & Fix, U. (Eds.): *Deutsch-deutsche Kommunikations-erfahrungen im arbeitsweltlichen Alltag*; pp. 113-152.
- Hellmann, M.W. (1997c) 'Sprach- und Kommunikationsprobleme in Deutschland Ost und West.' In: Schmirber, G. (Ed.): *Sprache im Gespräch – Zu Normen, Gebrauch und Wandel der deutschen Sprache*. München, Hanns-Seidel-Stiftung, pp. 53–87.
- Hellmann, M.W. (1999) 'Wörter in Texten der Wendezeit 1989/90 -- Ein Wörterbuch zur lexikographischen Erschließung des "Wendekorpus".' In: Jordanova, Ljubima (Hg.): *10 godini promjana v Iztotschna Evropa* (10 Jahre Wende in Osteuropa), (= Sociolinguistika Bd. 4). Sofia (Bulgarien) (BULLEKS), S. 11–39.
- Henne, H. (1995) (Kommentar) 'Hassen. Legendieren. Abschöpfen. Das Wörterbuch der Staatssicherheit'. In: ZGL; 23.2/95, S.210–214.
- Herberg, D. (1993) 'Die Sprache der Wendezeit als Forschungsgegenstand. Untersuchungen zur Sprachentwicklung 1989/90 am IDS [Bericht].' *Muttersprache* Bd. 103, H.3 [Themenheft *Sprache nach der Wende*], pp. 264–266.
- Herberg, D. (1994) 'Schlüsselwörter der Wendezeit. Ein Projekt zur Auswertung des IDS-"Wendekorpus"'. *SPRACHREPORT* (IDS), Nr. 1, p.4.
- Herberg, D. & Stickel, G. (1992) 'Gesamtdeutsche Korpusinitiative. Ein Dokumentationsprojekt zur Sprachentwicklung 1989/90' *Deutsche Sprache*, H. 2, pp. 185–192.
- Herberg, D., Steffens, D. & Tellenbach, E. (1997) *Schlüsselwörter der Wendezeit – Wörter-Buch zum öffentlichen Sprachgebrauch 1989/1990*. (= Schriften des Instituts für deutsche Sprache Bd. 6), Berlin, New York, de Gruyter.
- Heringer, H.J., Samson, G., Kauffmann, M. & Bader, W. (Eds.) (1994) *Tendenzen der deutschen Gegenwartssprache*. Tübingen, Niemeyer.
- Reiher, R. & Läser, R. (Eds.) (1993) *Wer spricht das wahre Deutsch? Erkundungen zur Sprache im vereinigten Deutschland*. Berlin, Aufbau Tb. Verlag.
- Reiher, R. (Ed.) (1995) *Sprache im Konflikt. Zur Rolle der Sprache in sozialen, politischen und militärischen Auseinandersetzungen*. Berlin, New York, de Gruyter (= Sprache Politik Öffentlichkeit Bd. 5).
- Samson, Gunhild (1994) 'Schlüsselwörter der Wende – "Sprachlosigkeit" und "Dialog".' In: Heringer, Samson, Kauffmann & Bader (Eds.) *Tendenzen der deutschen Gegenwartssprache*. Tübingen, Niemeyer, pp. 191–212.

- Schlosser, H.D. (1990) *Die deutsche Sprache in der DDR zwischen Stalinismus und Demokratie. Historische, politische und kommunikative Bedingungen*. Köln, Verlag Wissenschaft und Politik.
- Schlosser, H.D. (1992) 'Mentale und sprachliche *Interferenzen* beim Übergang der DDR von der Zentralplanwirtschaft zur Marktwirtschaft.' In: Welke, Sauer, Glück (Eds.) *Die deutsche Sprache nach der Wende* pp. 43-58.
- Schlosser, H.D. (1993) 'Die ins Leere befreite Sprache. Wende Texte zwischen Euphorie und bundesdeutscher Wirklichkeit.' *Muttersprache* Bd. 103, H.3, [Themenheft *Sprache nach der Wende*], pp. 219-230.
- Welke, K., Sauer, W.W. & Glück, H. (Eds.) (1992) *Die deutsche Sprache nach der Wende*. Hildesheim, Zürich, New York, Olms (= Germanistische Linguistik, 110-111).

*Reference works, bibliographies, documents*

- Ahrends, M. (1989) *Allseitig gefestigt. Stichwörter zum Sprachgebrauch der DDR*. (dtv Sachbuch Nr. 11126). München (dtv). [First edition (1986) entitled *Trabbi, Telespargel und Tränenpavillon*, München, Heyne Verlag]
- Cerný, J. et al. (Eds.) (1992) *Wer war wer – DDR. Ein biographisches Lexikon*. Berlin, Links-Verlag.
- DDR Handbuch (1985) *DDR Handbuch*. Edited by Bundesministerium für innerdeutsche Beziehungen, under the supervision of H. Zimmermann, 2 Vols, 3rd revised and expanded edition, Köln, Verlag Wissenschaft und Politik.
- Deutsche Bank AG – Zentrale Presseabteilung (Ed.) (1990) *Wirtschaftsbegriffe in Ost und West. 200 ausgewählte betriebswirtschaftliche Begriffe – interpretiert aus marktwirtschaftlicher beziehungsweise planwirtschaftlicher Sicht als Handlungs- und Orientierungshilfe*, Frankfurt/M.
- Edition Deutschland Archiv (Ed.) (1989) *Chronik der Ereignisse in der DDR*. Köln, Verlag Wissenschaft und Politik.
- Hellmann, M.W. (1992) *Wörter und Wortgebrauch in Ost und West. Ein rechnergestütztes Korpus-Wörterbuch zu Zeitungstexten aus den beiden deutschen Staaten. DIE WELT und Neues Deutschland 1949-1974*. 3 vols. (= Forschungsberichte des IDS Bd. 69.1-69.3) Tübingen, Narr.

- Hellmann, M.W., Melk, C. & Nikitopoulos, P. (to appear) *Alphabetisches Wörterbuch zum 'Wendekorpus' des IDS 1989/90*. Reihe Studien zur deutschen Sprache, Tübingen, Narr.
- Hellmann, M.W. (1999) *Wende-Bibliografie. Literatur und Nachschlagewerke zu Sprache und Kommunikation im geteilten und vereinigten Deutschland ab Januar 1990*. (= amades – Arbeitspapiere und Materialien zur deutschen Sprache, Mannheim, Institut für Deutsche Sprache, 1/1999).
- Kinne, M. & Strube-Edelmann, B. (1981) *Kleines Wörterbuch des DDR-Wortschatzes*. Düsseldorf, Schwann.
- Maier, G. (1991) *Die Wende in der DDR*. (= Reihe Kontrovers), 2nd updated edition. Bonn, Bundeszentrale für politische Bildung.
- Reiher, R. (Ed.) (1995) *Mit sozialistischen und anderen Grüßen. Porträt einer untergegangenen Republik in Alltagstexten*. Berlin, Aufbau Verlag.
- Weidenfeld, W. & Korte, R. (Eds.) (1996) *Handbuch zur deutschen Einheit*. Neuausgabe 1996. Bonn, Bundeszentrale für politische Bildung.
- Wimmer, M., Proske, C. et al. (Eds.) (1990) *'Wir sind das Volk!' Die DDR im Aufbruch. Eine Chronik in Dokumenten und Bildern*. (= Heyne Sachbuch. No. 19/113). München, Heyne.
- Wörterbuch der Staatssicherheit (1993): Der Bundesbeauftragte für die Unterlagen des Staatssicherheitsdienstes der ehemaligen Deutschen Demokratischen Republik (Ed.): *Das Wörterbuch der Staatssicherheit. Definitionen des MfS zur 'politisch-operativen Arbeit'*, Reihe A Dokumente, Nr. 1/93, Berlin.