

GernEdiT: A Graphical Tool for GermaNet Development

Verena Henrich

University of Tübingen
Tübingen, Germany.

verena.henrich@uni-
tuebingen.de

Erhard Hinrichs

University of Tübingen
Tübingen, Germany.

erhard.hinrichs@uni-
tuebingen.de

Abstract

GernEdiT (short for: GermaNet Editing Tool) offers a graphical interface for the lexicographers and developers of GermaNet to access and modify the underlying GermaNet resource. GermaNet is a lexical-semantic wordnet that is modeled after the Princeton WordNet for English. The traditional lexicographic development of GermaNet was error prone and time-consuming, mainly due to a complex underlying data format and no opportunity of automatic consistency checks. GernEdiT replaces the earlier development by a more user-friendly tool, which facilitates automatic checking of internal consistency and correctness of the linguistic resource. This paper presents all these core functionalities of GernEdiT along with details about its usage and usability.

1 Introduction

The main purpose of the GermaNet Editing Tool GernEdiT tool is to support lexicographers in accessing, modifying, and extending the GermaNet data (Kunze and Lemnitzer, 2002; Henrich and Hinrichs, 2010) in an easy and adaptive way and to aid in the navigation through the GermaNet word class hierarchies, so as to find the appropriate place in the hierarchy for new synsets (short for: synonymy set) and lexical units. GernEdiT replaces the traditional GermaNet development based on lexicographer files (Fellbaum, 1998) by a more user-friendly visual tool that supports versioning and collaborative annotation by several lexicographers working in parallel.

Furthermore, GernEdiT facilitates internal consistency of the GermaNet data such as appropriate linking of lexical units with synsets, connectedness of the synset graph, and automatic

closure among relations and their inverse counterparts.

All these functionalities along with the main aspects of GernEdiT's usage and usability are presented in this paper.

2 The Structure of GermaNet

GermaNet is a lexical-semantic wordnet that is modeled after the Princeton WordNet for English (Fellbaum, 1998). It covers the three word categories of adjectives, nouns, and verbs and partitions the lexical space into a set of concepts that are interlinked by semantic relations. A semantic concept is modeled by a *synset*. A synset is a set of words (called *lexical units*) where all the words are taken to have (almost) the same meaning. Thus a synset is a set-representation of the semantic relation of synonymy, which means that it consists of a list of lexical units.

There are two types of semantic relations in GermaNet: *conceptual* and *lexical relations*. Conceptual relations hold between two semantic concepts, i.e. synsets. They include relations such as hyperonymy, part-whole relations, entailment, or causation. GermaNet is hierarchically structured in terms of the hyperonymy relation. Lexical relations hold between two individual lexical units. Antonymy, a pair of opposites, is an example of a lexical relation.

3 The GermaNet Editing Tool

The GermaNet Editing Tool GernEdiT provides a graphical user interface, implemented as a Java Swing application, which primarily allows maintaining the GermaNet data in a user-friendly way. The editor represents an interface to a relational database, where all GermaNet data is stored from now on.

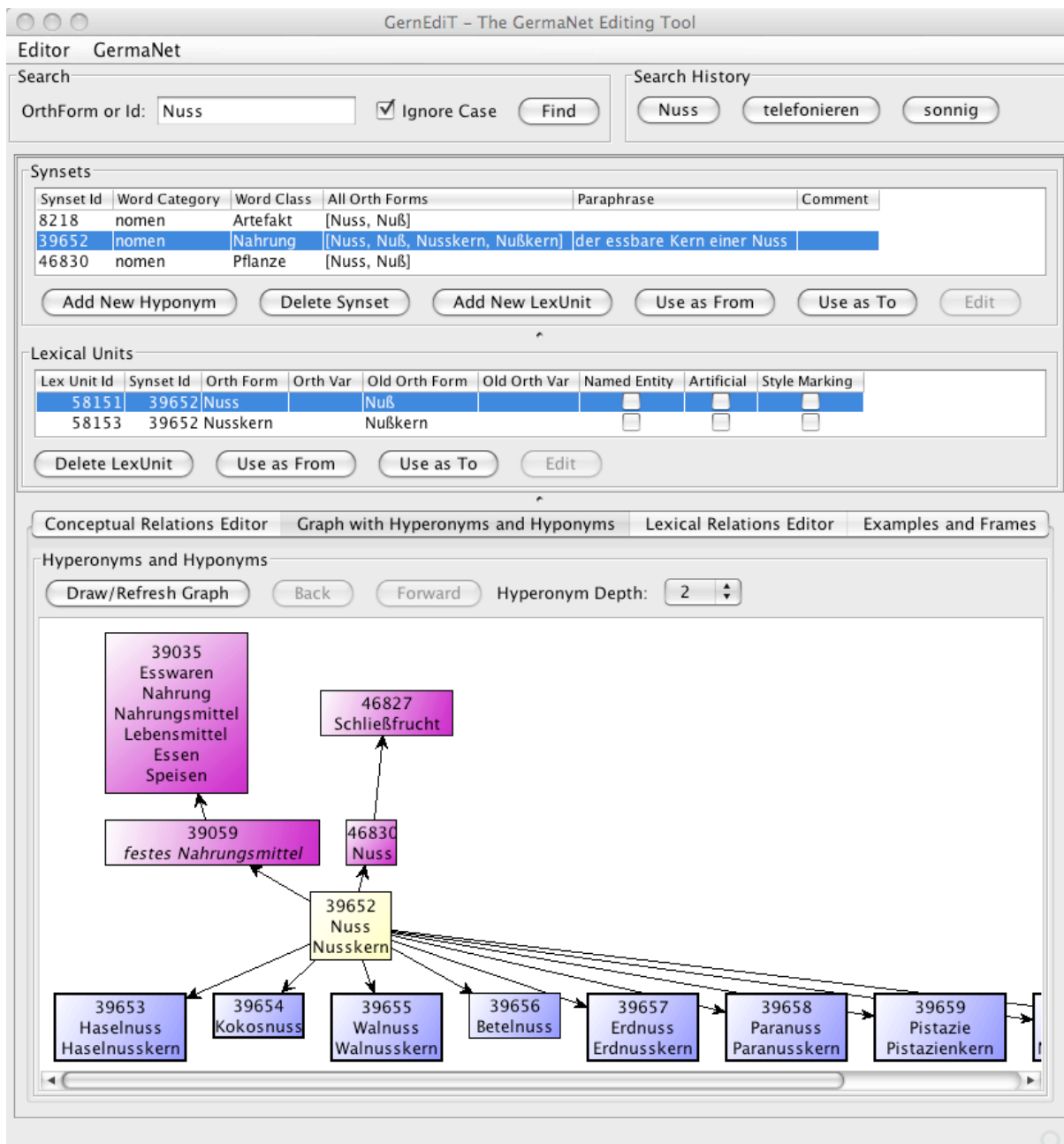


Figure 1. The main view of GernEdiT.

3.1 Motivation

The traditional lexicographic development of GermaNet was error prone and time-consuming, mainly due to a complex underlying data format and no opportunity of automatic consistency checks. This is exactly why GernEdiT was developed: It supports lexicographers who need to access, modify, and extend GermaNet data by providing these functions through simple button-clicks, searches, and form editing. There are several ways to search data and browse through the GermaNet graph. These functionalities allow lexicographers, among other things, to find the

appropriate place in the hierarchy for the insertion of new synsets and lexical units. Last but not least, GernEdiT facilitates internal consistency and correctness of the linguistic resource and supports versioning and collaborative annotation of GermaNet by several lexicographers working in parallel.

3.2 The Main User Interface

Figure 1 illustrates the main user panel of GernEdiT. It shows a *Search* panel above, two panels for *Synsets* and *Lexical Units* in the middle, and four tabs below: a *Conceptual Relations Editor*, a *Graph with Hyperonyms and Hyponyms*, a *Lexi-*

Lex Unit Id	Orth Form	Synset Id	Synonyms	Frames	Examples
84970	ausweiten	60447	[ausweiten]	NN.AN,NN.AN	"Wir werden unsere Bemühungen auf weitere Zielgruppen auswe..."
76139	abhalten	53955	[abhalten]	NN.AN.PP	"Wir versuchen, die Jugendlichen vom Rauchen abzuhalten."
84392	abspalten	60021	[abspalten]	NN.AN	"Wir müssen diese Eiweißmoleküle abspalten."
75212	aufwarten	53276	[aufwarten]	NN.DN	"Wichtige Politiker warteten dem König auf."
82681	aufleuchten	58722	[aufscheinen, aufleuchten]	NN	"Wenn das Warnlicht aufleuchtet, sollten Sie die Maschine sofort..."
74744	ausrasten	52956	[ausrasten, durchdrehen,...	NN	"Wegen der Nachricht rastete er aus."
76593	abtöten	54272	[devalisieren, abtöten]	NN.AN	"Vor der Wurzelbehandlung musste die Zahnärztin den Nerv abt..."
84660	ausrotten	60223	[ausrotten]	NN.AN	"Viele Tierarten sind durch Jagd und Industrie schon ausgerottet..."
74652	auseinanderdriften	52885	[auseinanderdriften]	NN	"Unsere Standpunkte drifteten zusehends auseinander."
76264	auswerten	54043	[auswerten]	NN.AN	"Sie werteten die neueste Marktanalyse aus."
75667	aufreten	53610	[auftreten, agieren]	NN.BM,NN.BR,NN.BL...	"Sie tritt selbstbewusst und entschieden auf.", "Sie tritt selbstbew..."
75876	aufreten	53772	[auftreten]	NN.PP.BM	"Sie tritt in der Öffentlichkeit sicher auf."
75415	abtreten	53424	[abtreten]	NN.PP	"Sie trat als Schauspielerin von der Bühne ab."
75515	abstreiten	53495	[ablehnen, abstreiten]	NN.AZ,NN.DS,NN.AN	"Sie stritt seine Anschuldigungen ab.", "Sie stritt seine Anschuldigu..."
85540	aufsplitten	60882	[splitten, aufsplitten]	NN.AN.PP,NN.AN,N...	"Sie splitteten das Vermögen in vier gleich große Teile auf.", "Sie..."
81894	antreten	58152	[antreten]	NN.BU	"Sie sind pünktlich zum Dienst angetreten."
79224	ausschelten	56189	[ausschimpfen, auszanke...	NN.AN	"Sie schalt ihren Sohn aus, nachdem er heimlich abgehauen war."
84227	ausschalten	59890	[abschalten, ausschalten,...	NN.AN,NN.AN	"Sie schaltete den Staubsauger ab.", "Er schaltete das Radio aus."
84221	anschalten	59887	[anstellen, anschalten, ein...	NN.AN	"Sie schaltete das Licht an."
78914	abraten	55945	[abraten]	NN.AZ,NN.DN,NN.DN.P...	"Sie riet ihm ab, nochmals dort anzurufen.", "Sie riet ihm ab, noch..."
77956	anarbeiten	55249	[anarbeiten]	NN.PP	"Sie muss gegen die neuen Richtlinien anarbeiten."
77111	abschalten	54617	[ausspannen, ausruhen, a...	NN	"Sie kann gut abschalten."
78039	auseinanderhalten	55311	[differenzieren, auseinan...	NN.AN	"Sie kann die Zwillingsschwestern nicht auseinanderhalten."
80008	aufheften	56784	[aufheften]	NN.AN.Pp	"Sie heftete das Tuch auf einer einfachen Plastikoberfläche auf."
75029	aufarbeiten	53141	[aufarbeiten]	NN.AN	"Sie hat viel Stoff aufzuarbeiten."
79575	anraten	56454	[anempfehlen, anraten]	NN.AN.DN	"Sie hatte ihm mehr Einfühlungsvermögen angeraten."
83768	abhärten	59546	[abhärten]	NN.AN.Pp,NN.AR.Pp...	"Sie hat ihren Körper durch gesunde Ernährung abgehärtet.", "Sie..."
78140	abtöten	55388	[abtöten, abwürgen]	NN.AN	"Sie hat ihre Gefühle für ihn abgetötet."

Figure 2: Filtered list of lexical units.

cal Relations Editor, and an Examples and Frames tab.

In Figure 1, a search for synsets consisting of lexical units with the word *Nuss* (German noun for: *nut*) has been executed. Accordingly, the *Synsets* panel displays the three resulting synsets that match the search item. The *Synset Id* is the unique database ID that unambiguously identifies a synset, and which can also be used to search for exactly that synset. *Word Category* specifies whether a synset is an adjective (*adj*), a noun (*nomen*), or a verb (*verben*), whereas *Word Class* classifies the synsets into semantic fields. The word class of the selected synset in Figure 1 is *Nahrung* (German noun for: *food*). The *Paraphrase* column contains a description of a synset, e.g., for the selected synset the paraphrase is: *der essbare Kern einer Nuss* (German phrase for: *the edible kernel of a nut*). The column *All Orth Forms* simply lists all orthographical variants of all its lexical units.

Which lexical units are listed in the *Lexical Units* panel depends on the selected synset in the *Synsets* panel. Here, *Lex Unit Id* and *Synset Id* again reflect the corresponding unique database IDs. *Orth Form* (short for: *orthographic form*) represents the correct spelling of a word according to the rules of the spelling reform *Neue Deutsche Rechtschreibung* (Rat für deutsche Rechtschreibung, 2006), a recently adopted spelling reform. In our example, the main orthographic form is *Nuss*. *Orth Var* may contain an

alternative spelling that is admissible according to the *Neue Deutsche Rechtschreibung*.¹ *Old Orth Form* represents the main orthographic form prior to the *Neue Deutsche Rechtschreibung*. This means that *Nuß* was the correct spelling instead of *Nuss* before the German spelling reform. *Old Orth Var* contains any accepted variant prior to the *Neue Deutsche Rechtschreibung*. The *Old Orth Var* field is filled only if it is no longer allowed in the new orthography.

The Boolean values *Named Entity*, *Artificial*, and *Style Marking* express further properties of a lexical unit, whether the lexical unit is a named entity, an artificial concept node, or a stylistic variant.

For both the lexical units and the synsets, there are two buttons *Use as From* and *Use as To*, which help to add new relations (see the explanation of Figure 3 in section 3.6 below which explains the creation of new relations).

3.3 Search Functionalities

It is possible to search for words or synset database IDs via the search panel (see Figure 1 at the top). The check box *Ignore Case* offers the possibility of searching without distinguishing between upper and lower case.

¹ An example of this kind is the German word *Delfin* (German noun for: *dolphin*). Apart from the main form *Delfin*, there is an orthographic variant *Delphin*.

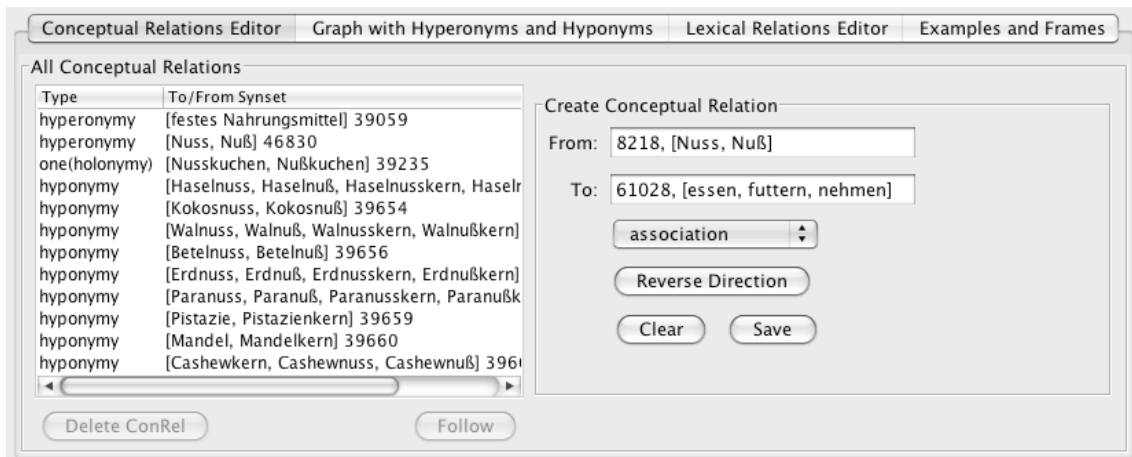


Figure 3. Conceptual Relations Editor tab.

Via the file menu, lists of all synsets or lexical units with their properties can be accessed. To these lists, very detailed filters can be applied: e.g., filtering the lexical units or synsets by parts of their orthographical forms. Figure 2 shows a list of lexical units to which a detailed filter has been applied: verbs have been chosen (see the chosen tab) whose orthographical forms start with an *a-* (see *starts with* check box and corresponding text field) and end with the suffix *-ten* (see *ends with* check box and corresponding text field). Only verbs that have a frame that contains *NN* are chosen (see *Frame contains* check box and corresponding text field). Furthermore, the resulting filtered list is sorted in descending order by their examples (see the little triangle in the *Examples* header of the result table). The number in the brackets behind the word category in the tab title indicates the count of the filtered lexical units (in this example 193 verbs pass the filter).

3.4 Visualization of the Graph Hierarchy

There is the possibility to display a graph with all hyperonyms and hyponyms of a selected synset. This is shown in the bottom half of Figure 1 in the tab *Graph with Hyperonyms and Hyponyms*. The graph in Figure 1 visualizes a part of the hierarchical structure of GermaNet centered around the synset containing *Nuss* and displays the hyperonyms and hyponyms of this synset up to a certain parameterized depth (in this case depth 2 has been chosen). The *Hyperonym Depth* chooser allows unfolding the graph to the top up to the preselected depth. As it is not possible to visualize the whole GermaNet contents at once, the graph can be seen as a window to GermaNet.

A click on any synset node within the graph, navigates to that synset. This functionality supports lexicographers especially in finding the appropriate place in the hierarchy for the insertion of new synsets.

3.5 Modifications of Existing Items

If the lexicographers' task is to modify existing synsets or lexical units, this is done by selecting a synset or lexical unit displayed in the *Synsets* and the *Lexical Units* panels shown in Figure 1. The properties of such selected items can be edited by a click in the corresponding table cell. For example by clicking in the cell *Orth Form* the spelling of a lexical unit can be corrected in case of an earlier typo was made.

If lexicographers want to edit examples, frames, conceptual, or lexical relations this is done by choosing the appropriate tab indicated at the bottom of Figure 1. By clicking one of these tabs, the corresponding panel appears below these tabs. In Figure 1 the panel for *Graph with Hyperonyms and Hyponyms* is displayed.

It is possible to edit the examples and frames associated with a lexical unit via the *Examples and Frames* tab. Frames specify the syntactic valence of a lexical unit. Each frame can have an associated example that indicates a possible usage of the lexical unit for that particular frame. The tab *Examples and Frames* is thus particularly geared towards the editing of verb entries. By clicking on the tab all examples and frames of a lexical unit are listed and can then be modified by choosing the appropriate editing buttons. For more information about these editing functions see Henrich and Hinrichs (2010).

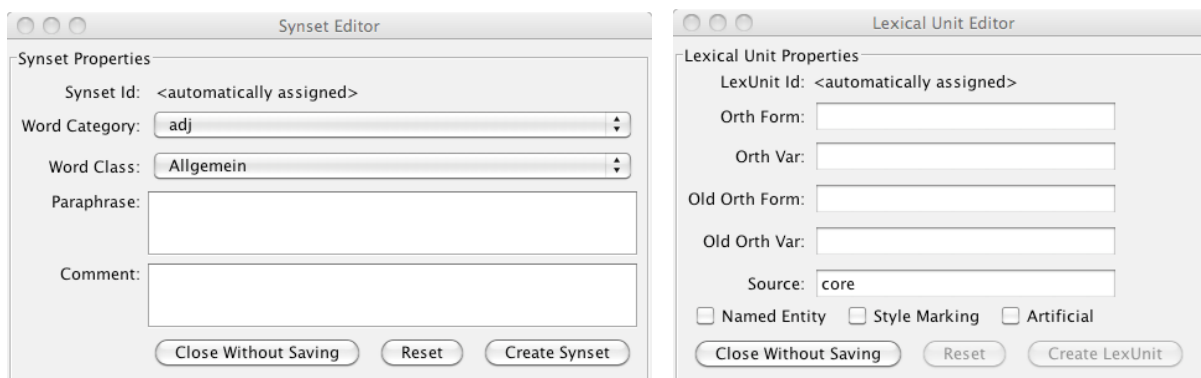


Figure 4. Synset Editor (left). Lexical Units Editor (right).

3.6 Editing of Relations

If lexicographers want to add new conceptual or lexical relations to a synset or a lexical unit this is done by clicking on the *Conceptual Relations Editor* or the *Lexical Relations Editor* shown in Figure 1.

Figure 3 shows the panel that appears if the *Conceptual Relations Editor* has been chosen for the synset containing *Nuss*. To create a new relation, the lexicographer needs to use the buttons *Use as From* and *Use as To* shown in Figure 1. This will insert the ID of the selected synsets from the *Synsets* panel in the corresponding *From* or *To* field in Figure 3. The button *Delete ConRel* allows deletion of a conceptual relation, if all consistency checks are passed.

The *Lexical Relations Editor* tab supports editing all lexical relations. It is not displayed separately for reasons of space, but it is analogue to the *Conceptual Relations Editor* tab for editing conceptual relations.

3.7 Adding Synsets and Lexical Units

The buttons *Add New Hyponym* and *Add New LexUnit* in the *Synsets* panel (see Figure 1) can be used to insert a new synset or lexical unit at the selected place in the GermaNet graph, and the buttons *Delete Synset* and *Delete LexUnit* remove the selected entry, respectively.

The *Synset Editor* in Figure 4 (on the left) shows the window which appears after a click on *Add New Hyponym*. When clicking on the button *Create Synset*, the *Lexical Unit Editor* (shown in Figure 4, right) pops up. This workflow forces the parallel creation of a lexical unit while creating a synset.

3.8 Consistency Checks

GernEiT facilitates internal consistency of the GermaNet data. This is achieved by the

workflow-oriented design of the editor. It is not possible to create a synset without creating a lexical unit in parallel (as described in section 3.7). Furthermore, it is not possible to insert a new synset without specifying the place in the GermaNet hierarchy where the new synset should be added. This is achieved by the button *Add New Hyponym* (see Figure 1) which forces the user to identify the appropriate hyperonym for the new synset to be added. Furthermore, it is not possible to insert a lexical unit without specifying the corresponding synset. On deletion of a synset, all corresponding data such as conceptual relations, lexical units with their lexical relations, frames, and examples, are deleted automatically.

Consistency checks also take effect for the table cell editing in the *Synsets* and *Lexical Units* panels of the main user interface (see Figure 1), e.g., the main orthographic form of a lexical unit may never be empty.

All buttons in GernEiT are enabled only if the corresponding functionalities meet the consistency requirements, e.g., if a synset consists only of one lexical unit, it is not possible to delete that lexical unit and thus the button *Delete LexUnit* is disabled. Also, if the deletion of a synset or a relation would violate the complete connectedness of the GermaNet graph, it is not possible to delete that synset.

3.9 Further Functionalities

There are further functionalities available through the file menu. Besides retrieving the up-to-date statistics of GermaNet, an editing history makes it possible to list all modifications on the GermaNet data, with the information about who made the change and how the modified item looked before.

GernEiT supports various export functionalities. For example, it is possible to export all GermaNet contents into XML files, which are used as an exchange format of GermaNet, or to

export a list of all verbs with their corresponding frames and examples.

4 Tool Evaluation

In order to assess the usefulness of GernEdiT, we conducted in depth interviews with the GermaNet lexicographers and with the senior researcher who oversees all lexicographic development. At the time of the interview all of these researchers had worked with the tool for about eight months. The present section summarizes the feedback about GernEdiT that was obtained in this way.

The initial learning curve for getting familiar with GernEdiT is considerably lower compared to the learning curve required for the traditional development based on lexicographer files. Moreover, the GermaNet development with GernEdiT is both more efficient and accurate compared to the traditional development along the following dimensions:

1. The menu-driven and graphics-based navigation through the GermaNet graph is much easier compared to finding the correct entry point in the purely text-based format of lexicographer files.
2. Lexicographers no longer need to learn the complex specification syntax of the lexicographer files. Thereby, syntax errors in the specification language – a frequent source of errors prior to development with GernEdiT – are entirely eliminated.
3. GernEdiT facilitates automatic checking of internal consistency and correctness of the GermaNet data such as appropriate linking of lexical units with synsets, connectedness of the synset graph, and automatic closure among relations and their inverse counterparts.
4. It is now even possible to perform further queries, which were not possible before, e.g., listing all hyponyms of a synset.
5. Especially for the senior researcher who is responsible for coordinating the GermaNet lexicographers, it is now much easier to trace back changes and to verify who was responsible for them.
6. The collaborative annotation by several lexicographers working in parallel is now easily possible and does not cause any management overhead as before.

In sum, the lexicographers of GermaNet gave very positive feedback about the use of GernEdiT and also made smaller suggestions for improving its user-friendliness further. This underscores the utility of GernEdiT from a practical point of view.

5 Conclusion and Future Work

In this paper we have described the functionality of GernEdiT. The extremely positive feedback of the GermaNet lexicographers underscores the practical benefits gained by using the GernEdiT tool in practice.

At the moment, GernEdiT is customized for maintaining the GermaNet data. In future work, we plan to adapt the tool so that it can be used with wordnets for other languages as well. This would mean that the wordnet data for a given language would have to be stored in a relational database and that the tool itself can handle the language specific data structures of the wordnet in question.

Acknowledgements

We would like to thank all GermaNet lexicographers for their willingness to experiment with GernEdiT and to be interviewed about their experiences with the tool.

Special thanks go to Reinhild Barkey for her valuable input on both the features and user-friendliness of GernEdiT and to Alexander Kislev for his contributions to the underlying database format.

References

- Claudia Kunze and Lothar Lemnitzer. 2002. *GermaNet – representation, visualization, application*. Proceedings of LREC 2002, Main Conference, Vol V. pp. 1485-1491, 2002.
- Christiane Fellbaum (ed.). 1998. *WordNet – An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- Verena Henrich and Erhard Hinrichs. 2010. *GernEdiT – The GermaNet Editing Tool*. Proceedings of LREC 2010, Main Conference, Valletta, Malta.
- Rat für deutsche Rechtschreibung (eds.) (2006). *Deutsche Rechtschreibung – Regeln und Wörterverzeichnis: Amtliche Regelung*. Gunter Narr Verlag Tübingen.