Enabling fast and correct typing in 'Leichte Sprache' (Easy Language)

Ina Steinmetz
Universität Koblenz-Landau
Computer Science Faculty
Universitätsstraße 1
56070 Koblenz
Germany
`inaschroeder@uni-koblenz.de`

Karin Harbusch
Universität Koblenz-Landau
Computer Science Faculty
Universitätsstraße 1
56070 Koblenz
Germany
`harbusch@uni-koblenz-de`

Abstract

Simplified languages are instruments for inclusion aiming to overcome language barriers. *Leichte Sprache* (LS), for instance, is a variety of German with reduced complexity (cf. Basic English). So far, LS is mainly provided for, but rarely written by, its target groups, e.g. people with cognitive impairments. One reason may be the lack of technical support during the process from message conceptualization to sentence realization. In the following, we present a system for assisted typing in LS whose accuracy and speed is largely due to the deployment of real time natural-language processing enabling efficient prediction and context-sensitive grammar support.

1   Motivation

Written language is essential to integration and autonomous participation in social and economic life (Bingel, 2018; Siegel and Lieske, 2014). An established instrument to overcome language barriers for people with cognitive impairment is *Leichte Sprache* (LS) (Maaß and Bredel, 2016; Bock, 2015; Maaß et al., 2014). This is a simplified variety of written German which follows a set of rules to reduce language complexity in order to simplify understandability (cf. Basic English).

In many languages, systems for automatic text simplification are available: e.g., see Bingel (2018) (English), Suter et al. (2016) (German) or Seretan (2012) (French). See the survey by Shardlow (2014) for more examples. However, few systems aiming to assist people who primarily rely on LS in correctly committing their thoughts in writing, deploy advanced natural-language processing-techniques. Many people with communication impairments use solutions for *Alternative and Augmentative Communication* (*AAC*) that combine symbols and words. For people with basic written German language skills, these solutions range from tools which only allow manual inflection (*MYCore13*) to systems featuring automatic inflection for simple subjects, different verb tenses, basic adjective declination and adaptive word prediction (*MindExpress4*, *Tobii Gateway*) and the construction of subordinate clauses (*Tobii Gateway*)[1].

The more complex the system, the more grammatical knowledge is required. For instance, the user is forced to select noun cases (*MindExpress4*) or to manually correct errors caused by incorrect predictions (*Tobii Gateway*). This inspired our research question: To which extent can natural-language processing by computer support users in producing coherent text in LS at the user's personal level of proficiency in spelling and clause construction?

In the following, we describe our prototype *EasyTalk* – as it can also read-out the typed texts – that supports fast and correct typing of complex content messages, with main emphasis on extensive use of linguistic processing and on interactive user guidance aiming to compensate for lack of grammatical knowledge and to ensure understandability and syntactic correctness.

2   *EasyTalk*: an ACC system for fast and syntactically correct typing in LS

Our system, called *EasyTalk*, uses a paraphrase generator (cf. Harbusch et al. (2014; 2007)) based on a lexicalized, unification-based Performance Grammar (Harbusch and Kempen, 2002; Kempen and

---

[1]MyCore13: `http://www.my-core.de/`, MindExpress4: `http://www.mindexpress.be/` and Tobii Gateway: `http://www.tobiidynavox.de/compass/gateway/`
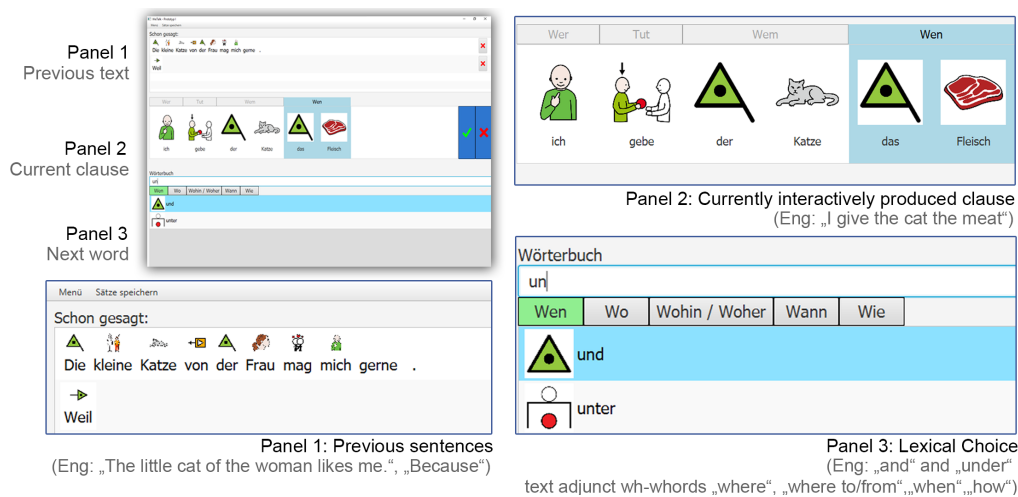
Figure 1: Top left: The overall organization of the three panels *EasyTalk* provides. Others: closeups.

Harbusch, 2002). It deploys a restricted set of grammar rules defined by the rule books of LS (Netzwerk Leichte Sprachel, 2013; Inclusion Europe, 2014; BITV 2.0, 2011) and their implicit linguistic effects (Maaß and Bredel, 2016). LS–rules such as *"Do not use commas"*, *"Avoid inversions"* and *"Use clear sentence structure"* are translated into main-clause patterns with a fixed subject-verb-argument-adjunct order. Genitive constructions (e.g. "the woman's cat") are avoided in favor of prepositional phrases ("the cat of the woman").

Figure 1 shows a snapshot of a text-production session with active elements highlighted in blue. Panel 1 contains the previously typed text: A statement followed by a cohesion marker. After each sentence the writer is reminded to select a cohesion marker from a set of conjunctions resembling RST connectors (Mann et al., 1992). These help users to write coherent texts. Panel 2 shows the currently produced sentence in the form of a skeleton of the clause structure in terms of headers for all obligatory grammatical functions couched in wh-words (cf. *wer* ('who'), *wem* ('whom'), *wen* ('what') according to the valency of the chosen verb *geben* ('give')). The finite verb is cued by *tut ('does')*. The blue area indicates the currently active position/inflection. To finish a sentence the user clicks the button with the green check mark. As a result, the sentence is displayed in panel 1. To discard the current sentence and start anew, the user clicks the blue button with the red X. Panel 3 displays a completion list of the currently typed prefix of the next word retrieved from the lexicon. The dictionary contains a list of all wordforms allowed in the currently constructed constituent. All suggestions are augmented by AAC symbols and the same RST-function cues used in panel 2. Further wh-cues of adjunct functions remind the user to specify the extensions explicitly. To write a sentence, the user selects words one-by-one from the suggestion list in panel 3. The linguistic context for the completion list is kept up to date at all times by our natural-language generator that yields the set of unified derivation trees. This way the user is always presented with a context-sensitive set of all wordforms that are syntactically correct choices. These choices can be further filtered by clicking on the cues mentioned above, thus recruiting additional syntactic support. If the word processing routine detects linguistic ambiguities, we temporarily replace panel 3 with a decision dialogue applying concepts of audience design (e.g. two cats mentioned in the text should be uniquely referred to or anaphorized, respectively) (Bell, 1984). The combination of interactive user guidance and inflected predictions provides reliable grammar support and speeds up typing.

A challenge with respect to the linear order of building a sentence is presented by complex verb constructions such as verb clusters including auxiliary or modal verbs. In order to satisfy the valency requirements of these verbs, our system asks the users if they want to add a second verb and updates the list of arguments accordingly. Striving for a correct and complete clause, we hereby knowingly urge the user to think in an unusual word order: In German, the arguments of the full verb often precede the verb, thus our system needs to know this verb in order to determine and support the formulation of its extensions.
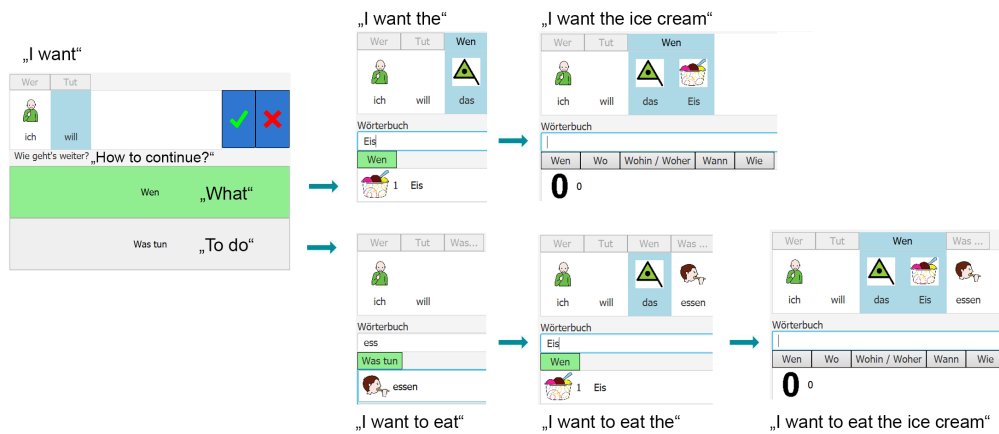
Figure 2: Snapshot of the users being presented with a decision dialogue, in which they may directly continue with the argument of the verb (top) or first choose a second verb (bottom).

Figure 2 illustrates the dialogue for handling complex verb constructions: As soon as the user chooses a verb that allows complex verb constructions – in the example the modal verb *will* ('want') – they are asked to decide whether to continue with the arguments of the first verb or to choose another verb. The grammar rules permit recursion here, i.e. the dialogue can produce a verb chain. As soon as the verb chain ends, the user fills the arguments in the usual manner as depicted in Figure 1. In Figure 2 this happens after the verb *essen* ('to eat') is selected. Our system automatically adjusts the position of the verb after each successively entered wordform of all possible arguments. By highlighting the current constituent the user is not distracted by the verb(s) at the end of the sentence.

3 Discussion and Conclusion

We have presented *EasyTalk*, a new AAC app for fast and correct LS typing which is largely based on linguistic processing by computer. We strive for correct and understandable text by providing a context-sensitively filtered list of wordforms and interactive user guidance tailored to minimize the grammatical knowledge needed to produce correct and coherent complex contents.

We performed preliminary tests aiming to mature the finally tested version: A group of five German learners at CEFR-level A2 explored the efficacy of the user interface and the linguistic writing support. Interviews with a group of five AAC-/LS-experts (consisting of one LS consumer/validator, two LS translators and two experts of AAC methods familiar with LS) enabled us to eliminate pitfalls in our system caused by too complex actions before testing with impaired users. Both groups were satisfied with the interactive grammar support and the customizable AAC symbols. Accordingly, we primarily got feedback on the syntactic correctness and the grammatical guidance by the German learners. The LS experts' concerns focused mainly on evaluating the usability of our system in day-to-day use and facilities for personalization and adaptations to special needs. As for the issue of verb clusters mentioned above, the language learners could easily manage. Learners coming from SVO languages even appreciated the decision dialogue to remind them of the SOV word order in German verb clusters. The expert group gave positive feedback, too: They did not voice any reservations about the verb-cluster dialogue while exploring the system. Elaborate testing with AAC users to achieve quantifiable results is projected.

As for future work and open problems, we want to elaborate user-customization: Not all users need LS at its utmost simplification. We plan facilities to tailor our rule set to the personalized skill level of the user. Moreover, the knowledge representation of typed text in our system (cf. panel 1) allows to transform the text into unrestricted German, so that the users can choose the language level of the expected reader. Our system canonically implements the rules of LS. However, these rules are still under discussion (Maaß and Bredel, 2016; Zurstrassen, 2017; Bock, 2019) and the textual implementation often differs from the rule books. Therefore we plan an in-depth analysis of a broad set of LS texts in order to offer the user optional grammar expansions of structures commonly used in LS texts.

References

Allan Bell. 1984. Language style as audience design. *Language in Society*, 13(2):145–204.

Joachim Bingel. 2018. *Personalized and Adaptive Text Simplification*. PhD Thesis. Department of Computer Science, Faculty of Science, University of Copenhagen, Denmark.

BITV 2.0. 2011. Verordnung zur Schaffung barrierefreier Informationstechnik nach dem Behindertengleich-stellungsgesetz (BarrierefreieInformationstechnik-Verordnung - BITV 2.0). *Barrierefreie-Informationstechnik-Verordnung vom 12. September 2011 (BGBl. I S. 1843), die zuletzt durch Artikel 1 der Verordnung vom 21Mai 2019 (BGBl. I S. 738) geändert worden ist.*

Bettina M. Bock. 2019. *"Leichte Sprache"–Kein Regelwerk: Sprachwissenschaftliche Ergebnisse und Praxisempfehlungen aus dem LeiSA-Projekt.(Vol. 5)* Frank & Timme, Berlin, Germany.

Bettina M. Bock. 2015. *Leichte Texte schreiben*. Frank & Timme, Berlin, Germany.

Karin Harbusch, Johannes Härtel, and Cristel-Joy Cameran. 2014. Automatic Feedback Generation for Grammatical Errors in German as Second Language Learners Focused on the Learner's Personal Acquisition Level. *Proceedings of EUROCALL 2014 - Call Design: Principles and Practice*, Groningen, The Netherlands.

Karin Harbusch, Camiel van Bruegel, Ulrich Koch and Gerard Kempen. 2007. Interarctive sentence combining and paraphrasing in support of integrated writing and grammar instruction: A new application area for natural sentence generators. *Proceedings of ENLG 2007 - 11th European Workshop on Natural Langauge Generation*, Dagstuhl, Germany.

Karin Harbusch and Gerard Kempen. 2002. A quantative model of word order and movement in English, Dutch and German complement constructions. *Proceedings of COLING 2002 - 19th International Conference on Computational Linguistics*, pp. 328–334, Taipei, Taiwan.

Inclusion Europe. 2014. *Informationen für alle - Europäische Regeln, wie man Informationen leicht lesbar und leicht verständlich macht* Inclusion Europe, Brussels, Belgium.

Gerard Kempen and Karin Harbusch. 2002. Performance Grammar: A declarative definition. Mariët Theune, Anton Nijholt, and Hendri Hondorp (eds.), *Computational Linguistics in the Netherlands 2001*, pp. 148–162, Rodopi, The Netherlands.

William C. Mann, Christian Matthias Ingemar Martin Matthiessen, and Sandra Annear Thompson. 1992. Rhetorical structure theory and text analysis. *Discourse description: Diverse linguistic analyses of a fund-raising text*:39–78.

Christiane Maaß and Ursula Bredel. 2016. *Leichte Sprache: Theoretische Grundlagen Orientierung für die Praxis*. Bibliographisches Institut, Berlin, Germany.

Christiane Maaß, Isabelle Rink, and Christiane Zehrer. 2014. Leichte Sprache in der Sprach- und Übersetzungswissenschaft. Susanne Johanna Jekat, Heike Elisabeth Jüngst, Klaus Schubert, Claudia Villiger (Hrsg.): *Sprache barrierefrei gestalten. Perspektiven aus der Angewandten Linguistik*, Frank & Timme, Berlin, Germany:53–86.

Melanie Siegel and Christian Lieske. 2014. Beitrag der Sprachtechnologie zur Barrierefreiheit: Unterstützung für Leichte Sprache. *Zeitschrift für Translationswissenschaft und Fachkommunikation*:40–78.

Netzwerk Leichte Sprache e.V. 2013. *Die Regeln für Leichte Sprache*. Netzwerk Leichte Sprache e.V., Berlin, Germany.

Violeta Seretan. 2012. Acquisition of syntactic simplification rules for French. *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, pp. 4019–4026, Istanbul, Turkey.

Matthew Shardlow. 2014. A survey of automated text simplification. *International Journal of Advanced Computer Science and Applications*, 4(1):58–70.

Julia Suter, Sarah Ebling, and Martin Volk. 2016. Rule-based automatic text simplification for German. *In Proceedings of the 13th Conference on Natural Language Processing (KONVENS 2016)*, pp. 279–287, Bochum, Germany.

Bettina Zurstrassen. 2017. Leichte Sprache–eine Sprache der Chancengleichheit. *Bettina M. Bock, Ulla Fix & Daisy Lange (Hg.), "Leichte Sprache" im Spiegel theoretischer und angewandter Forschung*, pp. 53–70. Frank & Timme, Berlin, Germany.