



# A Computer-Assisted Writing Tool for an Extended Variety of *Leichte Sprache* (Easy-to-Read German)

Karin Harbusch\* and Ina Steinmetz

*Institute for Computational Visualistics, Computer Science Faculty, Universität Koblenz-Landau, Koblenz, Germany*

*Leichte Sprache* (LS; easy-to-read German) defines a variety of German characterized by simplified syntactic constructions and a small vocabulary. It provides barrier-free information for a wide spectrum of people with cognitive impairments, learning difficulties, and/or a low level of literacy in the German language. The levels of difficulty of a range of syntactic constructions were systematically evaluated with LS readers as part of the recent *LeiSA* project (Bock, 2019). That study identified a number of constructions that were evaluated as being easy to comprehend but which fell beyond the definition of LS. We therefore want to broaden the scope of LS to include further constructions that LS readers can easily manage and that they might find useful for putting their thoughts into words. For constructions not considered in the *LeiSA* study, we performed a comparative treebank study of constructions attested to in a collection of 245 LS documents from a variety of sources. Employing the treebanks TüBa-D/S (also called VERBMOBIL) and TüBa-D/Z, we compared the frequency of such constructions in those texts with their incidence in spoken and written German sources produced without the explicit goal of facilitating comprehensibility. The resulting extension is called *Extended Leichte Sprache* (ELS). To date, text in LS has generally been produced by authors proficient in standard German. In order to enable text production by LS readers themselves, we developed a computational linguistic system, dubbed *ExtendedEasyTalk*. This system supports LS readers in formulating grammatically correct and semantically coherent texts covering constructions in ELS. This paper outlines the principal components: (1) a natural-language paraphrase generator that supports fast and correct text production while taking readership-design aspects into account, and (2) explicit coherence specifications based on *Rhetorical Structure Theory* (RST) to express the communicative function of sentences. The system's writing-workshop mode controls the options in (1) and (2). Mandatory questions generated by the system aim to teach the user when and how to consider audience-design concepts. Accordingly, users are trained in text production in a similar way to elementary school students, who also tend to omit audience-design cues. Importantly, we illustrate in this paper how to make the dialogues of these components intuitive and easy to use to avoid overtaxing the user. We also report the results of our evaluation of the software with different user groups.

**Keywords:** Augmentative and Alternative Communication (AAC), Controlled Languages (CL), Plain Language, Natural Language Generation (NLG), paraphrase generation, writing workshop/Schreibwerkstatt, *Leichte Sprache* corpus, treebank study

## OPEN ACCESS

### Edited by:

Scott A Hale,  
University of Oxford, United Kingdom

### Reviewed by:

Anouschka Foltz,  
University of Graz, Austria  
Hassan Mohebbi,  
European Knowledge Development  
Institute (EUROKD), Turkey

### \*Correspondence:

Karin Harbusch  
harbusch@uni-koblenz.de

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Communication

**Received:** 31 March 2021

**Accepted:** 30 September 2021

**Published:** 03 January 2022

### Citation:

Harbusch K and Steinmetz I (2022) A  
Computer-Assisted Writing Tool for an  
Extended Variety of *Leichte Sprache*  
(Easy-to-Read German).  
*Front. Commun.* 6:689009.  
doi: 10.3389/fcomm.2021.689009

## 1 INTRODUCTION

*Leichte Sprache* (LS) is a simplified variety of German. It was developed as part of the plain language movement of the 2000s (cf. easy-to-read English), which aimed to produce easy-to-understand texts for people with intellectual disabilities or learning difficulties (Bredel and Maaß, 2016, p. 60), who often have low literacy skills (Light et al., 2019). In Germany, LS is enshrined in law as the means of choice for providing accessible information in text form (BITV2.0, 2011).

The LeiSA project<sup>1</sup> identified a range of easily comprehensible syntactic constructions that are nonetheless beyond the scope of the core LS rules. It can be assumed that these constructions are used in language production, i.e., for putting thoughts into words. This leads to a research question concerning the *target grammar of our system*: What constructions might LS readers like to use in a writing tool? In order to obtain quantitative estimates of the incidence of the constructions evaluated in the LeiSA study, we built a parsed corpus of 245 published LS documents (a *Leichte Sprache treebank* we call LST). Constructions were found to have reasonable incidence, and no more than medium difficulty was included. The frequencies in LST of syntactic structures that had not yet been evaluated were compared to their frequencies in two treebanks of spoken and written German (VERBMOBIL and TüBa-D/Z, respectively), i.e., texts that had not been produced with the explicit goal of facilitating comprehensibility. The frequency of a construction's occurrence in the spoken corpus was compared with its frequency in the written one to provide the basis for an estimate of its ease of production; however, in order to keep the number of additional constructions to a minimum, we also judged whether or not a pure LS construction could easily replace one that is not included. The resulting extension of LS is called *Extended Leichte Sprache* (ELS).

To date, it has been usual for texts in LS to be produced by authors proficient in standard German and then evaluated for ease of comprehension by people with intellectual disabilities or learning difficulties (BITV2.0, 2011; Netzwerk Leichte Sprache, 2013). One factor preventing LS readers from producing texts themselves may be the lack of technical support during the process from *message conceptualization* (in the mind of the speaker/writer) to *sentence realization* (in a computer-assisted writing tool that remedies reading/writing deficits). Here, we consult terms used in *natural-language generation* (NLG) (cf. Section 3.1.2 for more details) to illustrate the complexity of the language production process involved in producing a text, compared to that of producing oral utterances in a face-to-face conversation.

To the best of our knowledge, there is no easy-to-use LS writing system that offers linguistic support beyond the phrasing of simple, partly personal sentences, let alone a system capable of teaching the concepts of written text production. The writing of coherent, understandable text requires an emphasis on audience-design concepts (Bell, 1984) because its target (i.e., the reader) cannot seek clarification—unlike the listener in face-to-face communication. German elementary school children learn

written text production by the widely applied method of the *Schreibwerkstatt/Schreibkonferenz* 'writing workshop' (see, e.g., Reichardt et al. (2014) for a broad survey). This technique is comparable to sentence-combining exercises in the Anglo-Saxon language area that teaches students to integrate sets of short, disconnected sentences to form longer, more effective ones (see Nordquist (2018) for an online introduction, Ney (1980) for the history, and Saddler and Preschern (2007) for the school context). This leads to two research questions concerning *assisted writing*: What individual support can help a range of users with intellectual disabilities, learning difficulties, and/or low literacy skills to write understandable, coherent text in ELS? Can we transform concepts from all stages in NLG into intuitive dialogues at the individual LS-reader level?

We present *ExtendedEasyTalk*, a writing tool with its main emphases on the extensive use of linguistic processing and on interactive user guidance aimed at compensating for a lack of grammatical knowledge and ensuring syntactic correctness and understandability. In order to produce a coherent text, *ExtendedEasyTalk* actively stimulates the user to add text-understandability and text-coherence elements, at both the constituent structure and the sentence-combining levels. For example, the sentences in (1-a) express the train of thought much better than the staccato phrases in (1-b), thanks to the use of coherence cues (*therefore/so, tomorrow, and the colon*).

- (1) a. Du bist krank. Darum: Ich will Dir helfen: Ich koche morgen das Mittagessen.  
You are ill. Therefore: I want you to help: I cook tomorrow the lunch.  
'You are ill, so I want to help you; I will cook lunch tomorrow.'
- b. Du bist krank. Ich helfe Dir. Ich koche Mittagessen.  
You are ill. I help you. I cook lunch.  
'You are ill. I will help you. I will cook lunch.'

This paper is organized as follows. In **Section 2**, we define ELS. After briefly introducing the core LS constructions, we list those evaluated in the LeiSA project (see **Section 2.2**). Then, we outline the comparative corpus study into constructions LS readers are likely to use in communication. (Readers wishing to skip the detailed linguistic argumentations can go directly to **Table 8**, which provides a list of the constructions included in ELS.) In **Section 3**, we present *ExtendedEasyTalk* as follows: First, we summarize the state-of-the-art technical writing support in the research area of *Augmentative and Alternative Communication* (AAC); a particular highlight of the descriptions is automatic NLG. Then, we give an intuitive overview of how the system works (see **Section 3.2**), before going into the technical details of sentence-constituent and sentence-coherence production (see **Sections 3.3 and 3.4**, respectively). We also outline the active control mechanism of *ExtendedEasyTalk* for teaching text-production concepts (see **Section 3.5**). Finally, we give the results of our evaluation with different user groups. **Section 4** draws some conclusions, discusses open issues, and suggests directions for future work.

## 2 EXTENDED LEICHTE SPRACHE

In this section, we define ELS, the target language of *ExtendedEasyTalk*. First, the syntactic constructions included in LS are outlined; this is followed by the evaluation results of

<sup>1</sup><https://research.uni-leipzig.de/leisa/> (2014–2018).

the LeiSA study. In **Section 2.3**, the treebank study is presented as follows: First, the creation of an LS treebank, LST, is described (see **Section 2.3.1**). Then, we quantify syntactic constructions in LST and compare them to standard German to classify their ease of comprehension. Finally, we give a summary of all additional constructions included in ELS, i.e., those that extend LS.

## 2.1 Leichte Sprache

Controlled/simplified natural languages, like Basic English (Ogden, 1930), have long been a topic of great interest (see Kuhn (2014) for a broad survey). The rules for Leichte Sprache were originally derived from practical experience (Bredel and Maaß, 2016, p. 60). The three main rule books (Netzwerk Leichte Sprache, 2013; Inclusion Europe, 2009; BITV2.0, 2011) have been the subject of previous scientific investigation (Lieske and Siegel, 2014; Maaß et al., 2014; Löffler, 2015; Zurstrassen, 2015; Bredel and Maaß, 2016; Bock, 2019; Nüssli, 2019; Pottmann, 2019; Hansen-Schirra and Maaß, 2020). Many rules concern the vocabulary (e.g., “Use easy words” or “No abbreviations”) or the avoidance of complex structures, for example, the use of:

- metaphors;
- more than one statement per sentence;
- punctuation other than: “.”, “?”, “!”, “:”;
- complex clauses;
- inversions;
- the genitive case;
- the passive voice;
- the subjunctive mood; and
- the simple past tense.

In other words, only main clauses are included in LS. In main declarative clauses, the canonical word order is subject–verb–object (SVO). All sentences should be phrased in the active voice, indicative mood, and present or present perfect tense.

The primary LS audience of people with cognitive impairments or learning difficulties is very heterogenous, and the available authentic text data by which to identify the range of constructions, LS readers naturally use in the process of formulating an idea are sparse. Usually, LS texts are written by authors proficient in standard German. Contrary to the recommendation in Netzwerk Leichte Sprache and Inclusion Europe, ease of comprehension is not always tested by members of the target readership.

Inspired by the finding of the LeiSA study that the majority of easily understandable LS texts do not strictly adhere to LS rules, we have explored possibilities for extending those rules to include the syntactic constructions that LS readers are likely to use when putting their thoughts into words.

## 2.2 Observations From the LeiSA Study

Part of the LeiSA study (Bock, 2019) was concerned with estimating the comprehension difficulty of individual syntactic constructions (see **Table 1**). Through well-established comprehension tests conducted using a five-point scale,

constructions—not only within but also beyond the scope of LS—were classified according to the error rates measured in experiments with participants with intellectual disabilities and functional illiteracy. We refer to the error ranges by the following difficulty levels:

- 0% ≤ error rate ≤ 5.9%: **easy**;
- 5.9% < error rate ≤ 12.5%: **low**;
- 12.5% < error rate ≤ 37.5%: **medium**;
- 37.5% < error rate ≤ 47.3%: **high**; and
- 47.3% < error rate ≤ 75.9%: **extreme**.

Constructions that show low error rates in the comprehensibility tests can be expected to be included in ELS. However, how can the full range of such constructions be obtained? We sought to identify the syntactic structures that LS readers are likely to use when putting their thoughts into words. In the following, we describe our search strategy, which leads to a broader set of rules, supported by the *ExtendedEasyTalk* system (see **Table 8** for a summary of the additionally included constructions).

## 2.3 A Comparative Treebank Study With Three German Corpora

Syntactic constructions that do not adhere to the LS rules are not hard to find in published LS documents. This suggests the advisability of inspecting a broad collection of LS texts and analyzing the constructions found therein. This requires the use of a Leichte Sprache treebank. As, to the best of our knowledge, no previous syntactically annotated corpus of LS texts exists, we have created one: LST (see **Section 2.3.1**).

Not all of the constructions we found in LS texts were evaluated with LS readers in the LeiSA project. In order to identify easy-to-understand constructions, we employed a treebank study as an alternative to the evaluation of example sentences by LS readers and compared the frequencies of constructions in LST to those in spoken and written standard German. We argue that constructions with high frequencies in spoken German are easy to produce due to the time-pressured nature of speech production. In a written text, the author is able to embellish the text in revision cycles, replacing simple constructions with more complex ones. Thus, the written corpus serves as a baseline. The appearance of a given construction in spoken language with a frequency higher than or equal to its appearance in written language is indicative of an easy-to-understand or unavoidable construction. Conversely, higher frequencies in written text imply difficult constructions employed under the non-time-critical conditions of revision and editing.

For the quantification of syntactic constructions in standard German, we used TüBa-D/S (also called VERBMOBIL), a treebank of spoken German, and TüBa-D/Z, a treebank of written German. To avoid confusion, we will use the name VERBMOBIL for TüBa-D/S. In the VERBMOBIL project (see, e.g., Stegmann et al., 2000 or Wahlster, 2000), more than 400 spontaneously produced spoken dialogues (concerning

**TABLE 1** | Overview of the syntactic constructions evaluated in the LeiSA study (Column 1: Difficulty level; Column 2: Is the concept included in the core rules of LS? (Y/N); Column 3: Characterization of the syntactic construction; Column 4: Examples from the LeiSA study; and Column 5: English translations).

Difficulty	LS	Linguistic phenomenon	LeiSA-study example	English translation
easy	N	Negations using <i>nicht</i> 'not'	<i>Der Bus fährt nicht.</i>	'The bus isn't running.'
easy	Y	Phrases and sentences with two or three elements	<i>ein kleiner Kürbis/das Handtuch falten</i>	'a small pumpkin/to fold the towel'
easy	Y	Prepositional phrases with the prepositions <i>in/auf</i> 'in/on'	<i>Die Schrauben sind in der Kiste./Alle Werkzeuge liegen auf dem Schrank.</i>	'The screws are in the box./All tools are on the cupboard.'
low	N	Temporal subordinate clauses with the conjunction <i>während/nachdem</i> 'while/after'	<i>Während Ihr Kind spielt, ist es zufrieden./Nachdem Sie alles markiert haben, sägen Sie die Form aus.</i>	'While your child is playing (s)he is happy./After you have marked everything, saw out the form.'
low	N	Subject relative clauses	<i>Mitarbeiter, die zufrieden sind, kommen gern zur Arbeit.</i>	'Employees, who are happy, enjoy coming to work.'
low	Y	Word forms in plural (they are hard to avoid)	no example provided	cf. <i>die Schrauben</i> 'the screws' mentioned above
low	Y	Spatial prepositions: <i>unter/über</i> 'under/above'	<i>den Brief unter das Buch legen/Das Handtuch hängt über dem Abfall-Eimer.</i>	'place the letter under the book/The towel hangs above the trash-can.'
medium	N	Passive forms with <i>werden</i> 'to be'—even reversible ones	<i>Der Antrag wird von Ihrem Chef gestellt./Dirk wird von Kristin gesucht.</i>	'The application is filed by your boss./Dirk is searched by Kristin.'
medium	Y	Present perfect tense	<i>Die Angestellten haben den Bus verpasst.</i>	'The employees have missed the bus.'
medium	N	Negations using <i>weder–noch</i> 'neither–nor'	<i>Weder der Tisch noch der Stuhl waren schön./Das Paket ist weder groß noch schwer.</i>	'Neither the table nor the chair were beautiful./The package is neither big nor heavy.'
medium	Y	Double object constructions	<i>Die Küchenhilfe zeigt dem Kollegen den Speiseplan./Das Amt schickt Ihnen einen Brief.</i>	'The kitchen aid shows the colleague the menu./The government agency sends you a letter.'
medium	N	Coordination with 'and' and elliptical structures	<i>Der Schuh ist auf dem Stift und ist blau.</i>	'The shoe is on the pencil and is blue.'
high	N	Topicalization	<i>Tische und Stühle müssen Sie zuerst wegräumen. Bei Verbrennungen helfen diese Maßnahmen.</i>	'Tables and chairs, you have to remove first./In case of burns, these measures help.'
high	N	<i>dass</i> 'that' clauses with varying reference of the accusative object	<i>Der Junge sieht, dass die Frau sich/sie sieht.</i>	'The boy notices that the woman sees herself/(her/ them).'
extreme	N	Non-subject relative clauses	<i>Das Buch, auf dem der Stift ist, ist rot.</i>	'The book, on which the pencil is, is red.'

**TABLE 2** | Overview of the three treebanks.

	LST	VERBMOBIL	TüBa-D/Z
Number of corpus graphs/syntactic structures	29,170	38,328	104,785
Number of tokens/leaves of structures	255,714	360,084	1,959,038
Average number of tokens per corpus graph	8.8	9.4	18.7
Number of inner nodes	457,324	496,466	2,402,421
Number of edges	683,868	818,222	4,353,888

appointment scheduling) were transliterated and syntactically annotated. The Tübinger Baumbank des Deutschen/Zeitungskorpus (TüBa-D/Z; see, e.g., Telljohann et al., 2009) is a syntactically annotated corpus based on the German newspaper *die Tageszeitung (taz)*.

**Table 2** shows the overall sizes of the three investigated corpora. In all three corpora, corpus graphs (i.e., depictions of the syntactic structures) do not necessarily encode complete sentences in the linguistic sense—they also include, for example, headlines, terms in brackets, incomplete turns, and self-repairs. Tokens (i.e., the leaves of corpus graphs) cover not only word forms but also punctuation. As expected, in LST, the average number of tokens per corpus graph (roughly speaking, the sentence length) is shorter than in spoken utterances, although not to a great degree. This surprising

circumstance results from long item lists (cf. example (8) below) that do not occur in spoken utterances, and the fact that we added missing punctuation symbols to improve the automatic syntactic analysis (*parsing*).

In the next section, we introduce the *Leichte Sprache* treebank LST used in our study.

### 2.3.1 Building the *Leichte Sprache* Treebank LST, a Syntactically Annotated *Leichte Sprache* Corpus

From a variety of sources freely available on the internet, spanning the years 2018–2021, we assembled a corpus of 245 LS texts with more than 300,000 word forms. To build a representative data set of LS texts of sufficient variety, we selected a broad spectrum of institutions, authors, and validators: according to the credits, at least 153 authors, 116

validators, and 53 institutions were involved in the creation of these texts. Each corpus graph in LST provides a feature with detailed source information of the original text. (For this purpose, we adopt the practice followed in historical corpora; see, e.g., the treebank of Old High German *Tatian*; Petrova et al., 2009.) We had originally planned to use this information to distinguish the following two subcorpora:

- LST-WithP, comprising only those texts that were proofread by LS readers, and
- LST-NoP, including texts without explicitly mentioned LS-reader participation.

To our surprise, the two subcorpora do not differ with respect to the number of violations of LS constraints; we therefore omitted the planned step of investigating differences between their construction frequencies. Nevertheless, we identified that a majority of the texts follow the LS rules. These texts not only deal with simple topics (e.g., fairytales) but also concern many spheres of life, including patient decrees, voters' rights, financial matters, and laws of succession. The implication is that the conformity of a text to LS rules does not depend on the complexity (or simplicity) of the topic but on whether or not its authors are aware of *best practices* (e.g., rephrasing *if* sentences as questions). We plan to investigate this observation in more detail in the future.

In the following, we sketch the process of obtaining the syntactic structures by parsing.

Preprocessing was done to improve the parsing results. From all PDF files, the plain text was extracted. In the extracted text, meta-text (e.g., running titles, page numbers, tables of contents, address lists, or links) was removed. Moreover, mediopoints (a specific functional LS symbol to segment compound nouns) and dashes without capitalized trailing word forms were removed to make full use of the compound analysis during parsing. In a series of pretests, we noticed that line breaks cause underspecification (cf. subscript “Dat./-” in (2-a), i.e., dative-case assignment in the sentential context vs. morphological underspecification if parsed in isolation). More seriously, it is not unusual that sentence fragments in separate lines are parsed incorrectly (cf. (2-b) where the isolated second line denotes a finite main declarative clause due to matching subject-verb agreement). We therefore removed line breaks (represented by the symbol “//” in the following) and colons within clauses (cf. (2-c)) in order to obtain correct case or grammatical function assignments.

- (2) a. Ein Impfstoff schützt vor einer // Krankheit.  
A vaccine protects against an illness<sub>DAT/-</sub>
- b. Sie dürfen zum Beispiel nicht in den Urlaub fahren // oder Freunde besuchen.  
You may for example not into the holiday drive or friends<sub>ACC/NOM</sub> visit<sub>INF/FIN</sub>  
'For example, you may not go on vacation or visit friends.'
- c. Die Person heißt: // der Bevollmächtigte.  
The person is<sub>called</sub> the authorized<sub>representative</sub><sub>PRED/NOM/-</sub>

Although subordinate clauses are not included in LS, they occur in the corpus material, with or without the correct German punctuation but usually in a separate line. We reconstructed the overall sentence, including punctuation, according to conjunction and the verb position. In German declarative clauses with at least three constituents, main and subordinate clauses differ in word

ordering. The finite verb form fills the second constituent position in main clauses (V2), whereas it is final in subordinate clauses (VF). For example, relative clauses, especially those where the relativizer differs from *der/die/das*<sub>inflected</sub> ‘who/which,’ are not recognized as such when parsed in isolation (PRELS refers to a relative pronoun; PWS to a substituting interrogative pronoun):

- (3) Dann berichtet sie in ihrer Fernsehsendung genau das, // (was<sub>PRELS/PWS</sub> die Menschen erzählt haben) //  
Then reports she in her tv\_show exactly that what the people told have  
erzählt haben) *relative-clause/root-clause*  
'Then she reports in her tv show exactly what the people have said.'

However, we did not change the typical LS construction where a conjunction/causative adverb stands in a separate line—possibly followed by a colon—when the next sentence has V2 word order, as in example (4):

- (4) Das ist nicht gut. // Denn: // Die Bauern müssen<sub>VMFIN</sub> deshalb sehr viele Tiere halten<sub>VVINFIN</sub>.  
This is not good because the farmers have<sub>to</sub> therefore very many animals keep  
'This is not good because the farmers have therefore to keep a great many animals.'

In example (5), the word order of the second line would be parsed in isolation as a yes/no question:

- (5) Damit Sie das Heft gut lesen können: // Haben wir immer die männliche Form geschrieben.  
So<sub>that</sub> you the leaflet good read can have we always the male form written  
'So that you can read the leaflet easily we use the male gender form exclusively.'

In example (6), an obligatory/complement clause, unrelated to the main clause, would remain:

- (6) Familien-Hebammen achten darauf: // Dass es Ihrem Kind immer gut geht. // Und dass es Ihnen immer gut geht.  
Family-midwives make\_sure that your baby is always well goes and that it you always well goes  
'Family midwives make sure that your baby is always fine and that you are always fine.'

Lists are very frequent constructions in LS texts. For correct parsing, it is necessary to revise the punctuation throughout the sentence. The list in (8) cuts into the main clause, although the clause in (7) ends with a colon. The list starts with a prepositional phrase modifying the direct object of the main clause. Then, it continues with a long list of subordinate clauses. The reconstructed complex sentence consists of more than 75 words. This example explains the unexpectedly high average length of the corpus graphs in LST.

- (7) Das Robert-Koch-Institut sammelt Informationen:  
The Robert Koch Institute collects information
- (8) a. - über das Corona-Virus  
about the Corona virus
- b. - Wie viele Menschen krank sind  
how many people ill are  
'how many people are ill'
- c. - Wo Menschen krank sind  
where people ill are  
'where people are ill'
- d. - Wie viele Menschen wieder gesund sind  
how many people again healthy are  
'how many people have recovered'
- e. - Ob es besser oder schlechter wird  
whether it better or worse gets  
'whether it gets better or worse'
- ...

In a further change, we systematically added question marks to make questions more easily recognizable in LST. Although they are included in LS, a large majority of questions in the corpus are printed with a full stop or without any punctuation.

**Quality of the syntactic structures.** A small LS sample text was evaluated with different parsers (e.g., from the wide spectrum of deep-learning approaches; see Linzen and Baroni (2021) for a recent survey). We finally decided to employ PARZU (Sennrich et al., 2013). PARZU is a dependency parser for German with a rich part-of-speech (POS) and morphological tag inventory (see Tritscher (2016) for an evaluation with German prose text). In a random sample of 100 sentences from the overall LS corpus, we identified four, partially very minor, errors (provided in the **Supplementary Material**). Consequently, the overall quality of LST is shown to be less perfect than the manually inspected standard German treebanks; however, given the overall treebank size of LST, the accuracy is deemed sufficient for the identification of clear trends. Thus, using the power of a treebank search that exceeds the scope of a manual inspection of a small sample or a pure word-form-based text search, LST gives rise to valuable insights.

In our study, the dependency trees produced by PARZU were transformed into the TIGER-XML format (König and Lezius, 2003), in which VERBMOBIL and TüBa-D/Z are also available. All three treebanks were inspected with TIGERSearch<sup>2</sup> (König and Lezius, 2003). Note that VERBMOBIL is not morphologically annotated. Therefore, some queries cannot be answered in this corpus (they will be referred to as “n.a.”, i.e., not applicable, in the tables).

### 2.3.2 Typical Syntactic Constructions in LST Compared to VERBMOBIL and TüBa-D/Z

In the following section, we study the frequencies of a wide range of syntactic constructions in LST. Constructions that fulfill one of the following conditions are added to ELS:

- Constructions that have at the most medium-level difficulty for LS readers (according to the LeiSA study) and that occur frequently in LST; and
- Constructions not covered by the LeiSA study, whose LST frequencies compare favorably to their frequencies in spoken German, but which cannot be easily transformed into pure LS constructions.

We present the syntactic phenomena, ordered according to their level of construction complexity: (1) word-related, (2) phrase-related, and (3) clause type-related constructions. Within each level, we first refer to the phenomena mentioned in **Table 1**; then, we discuss typical simple constructions that are beyond the scope of pure LS and were not evaluated in the LeiSA study. The systematic nature of this search demonstrates that we assessed the whole range of simple syntactic constructions. For reasons of space, we omit many details here, especially when no constructions are added to ELS.

**Negations.** **Table 3** provides the frequencies of several negation words. The absolute numbers are provided in

brackets. (This format is preserved in **Tables 4–7**.) Negation is forbidden in LS; however, it is difficult to avoid completely (Bock, 2017). As *nicht* ‘not’ is easy for LS readers to understand (according to the LeiSA study), we opt to add it to ELS. According to the frequency of its occurrence in LST (comparable to that in VERBMOBIL), *kein<sub>inflected</sub>* should also be added to prevent forcing a reformulation with *nicht*. All other constructions including negation are very infrequent in both LST and VERBMOBIL.

**Prepositions.** The POS of a preposition is distinguished in pre-positioned prepositions (APPRs) (POS starting with APPR = APPR.\*), post-positioned prepositions (APPOs), and partly fronted and partly trailing prepositions (circum-positioned POS = APZR [according to the German term *Zirkumposition*]). APPRART specifies prepositions with an agglutinated definite article (APPR.\* refers to APPR+APPRART).

APPRART should not occur in LS; nevertheless, this construction is frequently used. Half of the cases occur in the idiomatic prepositional phrase (PP) *zum Beispiel* ‘for example,’ which should not be abbreviated in LS. In practice, the use of APPRART makes sentences shorter. Moreover, using *zu dem Beispiel* for z.B. ‘e.g.’ sounds odd in German. We therefore add APPRART to ELS. **Table 4** provides frequencies for the explicitly mentioned prepositions in **Table 1**. Note that the treebanks do not distinguish spatial/temporal use. Thus, the frequencies presented here for the prepositions *unter/über* are overall figures.

The two partially or completely post-positioned preposition types, APPO and APRZ, occur in LST. All 47 cases of APPOs specify a temporal duration with *lang* ‘long’, as in *10 Tage lang<sub>APPO</sub>* ‘for 10 days’. There are 16 APZRs occurring (e.g., *von zuhause aus<sub>APZR</sub>* ‘from home’ and *von Anfang an<sub>APZR</sub>* ‘since the beginning’). Under LS rules, all but genitive-taking prepositions (absent in LST) are included; therefore, no further extension is suggested for prepositions.

**Nouns and pronouns.** In noun phrases (NPs), the following POS tags occur as heads: NE (proper noun); NN (lexical noun); and PRO (pronoun), which summarizes PDS (substituting demonstrative pronoun), PIS (substituting indefinite pronoun), PPOSS (substituting possessive pronoun), PRELS (relative pronoun), PRF (reflexive personal pronoun), PPER (irreflexive personal pronoun), and PWS (substituting interrogative pronoun). In **Table 5**, N.\*+PRO refers to any NP head filler, where POS = N.\* refers to NE+NN (suppressed in the table because it can be reconstructed by subtracting the frequency of PRO from N.\*+PRO; e.g., in LST,  $75.51 = 1 - 25.49\%$ , and  $58,832 = 78,961 - 20,129$  cases). We first investigate the morphological features *number = plural* and *case = genitive*, which were studied in the LeiSA project, in LST, and TüBa-D/Z. Then, we present the frequencies of the individual PRO types in all three treebanks. The frequencies of lexical nouns are given in the bottom row. The LeiSA study excluded pronouns and proper nouns from its investigations.

Plural forms are of low difficulty. They occur slightly more frequently in LS than in TüBa-D/Z (no data are available for VERBMOBIL). This construction is included in LS. The frequency of genitives, forbidden in LS, is very low; therefore, we do not add genitives to ELS, although we note that such

<sup>2</sup>To replicate the data, a TUNDRA-based search (see <https://weblicht.sfs.uni-tuebingen.de/Tundra/>) for all three specified corpora in the dependency-tree format offers another option. The query format in TUNDRA is based on TIGERSearch. All queries used in the following are provided in the **Supplementary Material**.

**TABLE 3** | Frequencies of negations in the three treebanks.

	LST		VERBMOBIL		TüBa-D/Z	
Freq. of <i>nicht</i> 'not' in all tokens (ats)	6.31 %	(1,840)	6.29 %	(2,411)	12.84 %	(13,457)
Freq. of <i>nichts</i> 'nothing' in ats	0.27 %	(78)	0.55 %	(209)	1.14 %	(1,190)
Freq. of <i>nie(mals)</i> 'never' in ats	0.17 %	(50)	0.08 %	(32)	0.64 %	(674)
Freq. of <i>niemand</i> <sup>inflected</sup> 'nobody' in ats	0.39 %	(114)	0.01 %	(3)	0.40 %	(417)
Freq. of <i>kein</i> <sup>inflected</sup> 'no' in ats	2.19 %	(640)	1.59 %	(611)	3.80 %	(3,982)
Freq. of <i>weder – noch</i> 'neither-nor' in ats	0.00 %	(0)	0.01 %	(4)	0.28 %	(289)

**TABLE 4** | Frequencies of APPR.\* in the three treebanks.

	LST		VERBMOBIL		TüBa-D/Z	
Freq. of APPR.* in all tokens	8.09 %	(20,690)	6.55 %	(23,588)	8.86 %	(173,574)
Freq. of APPRART in APPR.*	1.72 %	(4,401)	1.79 %	(6,446)	1.53 %	(29,944)
Freq. of <i>in/auf</i> in APPR.*	2.49 %	(6,359)	1.60 %	(5,755)	2.78 %	(54,464)
Freq. of <i>unter/über</i> in APPR.*	0.02 %	(54)	0.01 %	(22)	0.34 %	(6,752)

**TABLE 5** | Frequencies of (pro)nouns in the three treebanks.

	LST		VERBMOBIL		TüBa-D/Z	
<b>Total number of N.*+PRO (n+p)</b>		78,961		94,725		579,511
Freq. of N.*+PRO in plural in n+p	27.71 %	(21,883)	n.a		24.17 %	(140,089)
Freq. of genive case-n+ps in n+p	0.68 %	(533)	n.a		9.14 %	(52,953)
Freq. of PRO in n+p	25.49 %	(20,129)	50.79 %	(48,108)	17.93 %	(103,897)
Freq. of PDS in PRO	10.08 %	(2,030)	18.59 %	(8,944)	8.13 %	(8,445)
Freq. of PIS in PRO	15.25 %	(3,070)	5.79 %	(2,786)	15.01 %	(15,598)
Freq. of PPOSS in PRO	0.00 %	(0)	0.04 %	(18)	0.03 %	(35)
Freq. of PRELS in PRO	3.64 %	(733)	0.75 %	(361)	14.78 %	(15,359)
Freq. of PRF in PRO	7.94 %	(1,599)	5.75 %	(2,764)	14.15 %	(14,701)
Freq. of PPER in PRO	59.47 %	(11,971)	67.03 %	(32,247)	44.94 %	(46,695)
Freq. of PWS in PRO	3.61 %	(726)	2.05 %	(988)	2.95 %	(3,064)
Freq. of NN in NN+NE	91.85 %	(54,036)	87.11 %	(40,610)	78.41 %	(372,949)

**TABLE 6** | Frequencies of verb forms in the three treebanks.

	LST		VERBMOBIL		TüBa-D/Z	
<b>Total number of verb forms (V.*)</b>		41,996		50,679		243,705
Freq. of VA.* in V.*	22.92 %	(9,625)	35.78 %	(18,134)	28.60 %	(69,706)
Freq. of VM.* in V.*	19.33 %	(8,117)	13.68 %	(6,934)	22.60 %	(19,983)
Freq. of V.FIN or V.IMP in V.*	71.88 %	(30,186)	77.26 %	(39,154)	66.90 %	(163,033)
Freq. of V.PP in V.*	6.28 %	(2,639)	2.76 %	(1,400)	16.35 %	(39,852)
Freq. of V.INF/V.IZU in V.*	21.84 %	(9,171)	19.98 %	(10,135)	16.75 %	(40,820)
Freq. of preterite tense in V.*	2.16 %	(907)	n.a		20.13 %	(49,049)
Freq. of passives in V.*	1.51 %	(633)	0.11 %	(53)	5.43 %	(13,238)
Freq. of subjunctive mood in V.*	0.73 %	(308)	n.a		6.06 %	(14,744)
Freq. of double object verbs in V.*	1.08 %	(454)	2.09 %	(1,057)	1.74 %	(4,240)

constructions are often used in idiomatic expressions, e.g., for the names of institutions. A comparison of pronoun frequencies places LST somewhere between the written and spoken corpora. However, we did not expect high numbers of pronouns, for the following reasons: The three major LS rule

sets Netzwerk Leichte Sprache, 2013, Inclusion Europe, 2009, and BITV2.0 (2011) insist on consistent naming, i.e., using exactly the same word for the same thing/person throughout a text; additionally, the LS rule set of Inclusion Europe forbids pronominal resumption, favoring nominal resumption. In

**TABLE 7** | Sentence complexity in the three treebanks.

	LST		VERBMOBIL		TüBa-D/Z	
<b>Total of sent.-onset frontfieds (sofs)</b>		29,170		11,600		72,586
Freq. of NP-SB in sofs	47.48 %	(13,851)	48.85 %	(5,667)	50.48 %	(36,645)
Freq. of short NP-SB in SB sofs	91.31 %	(12,648)	98.61 %	(5,588)	34.62 %	(25,128)
<b>Total of relative clauses (rels)</b>		657		427		17,017
Freq. of short rels in rels	86.61 %	(569)	90.16 %	(385)	56.47 %	(9,609)
Freq. of SB-rels in rels	64.08 %	(421)	n.a		64.51 %	(10,978)
Freq. of short SB-rels in SB rels	86.94 %	(366)	n.a		57.03 %	(6,261)
<b>Total of subordinate clauses (subs)</b>		2,811		3,294		23,302
Freq. of short subs in subs	76.91 %	(2,161)	80.90 %	(2,665)	47.94 %	(11,171)
Freq. of 'while/after' (wcls) in subs	0.18 %	(5)	0.18 %	(6)	9.99 %	(2,327)
Freq. of short wcls in wcls	40.00 %	(2)	83.33 %	(5)	35.45 %	(825)
Freq. of 'if'cls. (icls) in wcls	53.90 %	(1,515)	37.74 %	(1,243)	15.52 %	(3,616)
Freq. of short icls in icls	79.87 %	(1,210)	88.74 %	(1,103)	61.37 %	(2,219)
Freq. of 'that' cls. (tcls) in wcls	16.22 %	(456)	26.29 %	(866)	34.46 %	(8,031)
Freq. of short tcls in tcls	72.81 %	(332)	71.25 %	(617)	37.42 %	(3,005)

particular, many occurrences of PIS that characterize abstract referents (e.g., *man* 'one', *jemand*<sub>inflected</sub> 'somebody', *etwas* 'something', and *alle*<sub>inflected</sub> 'all') resemble those in written text. Due to their frequency, to shorten the resulting sentences, and to enable the use of abstract referents to circumvent passive constructions, we include all pronouns in ELS. The frequency of proper nouns in LST is similar to that in VERBMOBIL. Because this construction is already included in LS, no extension is required.

**Verb forms.** Before examining phrases and clauses, we study verb forms (see **Table 6**). Part-of-speech tags starting with "VA" (POS = VA.\*) refer to auxiliary verbs, including the copula use of 'to be.' The prefix "VM" (POS = VM.\*) characterizes modal verbs, and "VV" (POS = VV.\*) lexical verbs. The total VV.\* frequencies are omitted in favor of reconstructing  $VV.* = V.* - VA.* - VM.*$  for each corpus. The frequencies, resembling those of TüBa-D/Z, qualify LST as a variety of written text. Not distinguishing between the verb types (V. = VA + VM + VV), the proportions of finite verb forms (V.FIN/V.IMP, i.e., any POS type ending with a finite [FIN] or imperative [IMP]) in declarative, interrogative, and imperative clauses give a rough estimation of clause simplicity. A clause with a finite lexical verb form that is not one of the few non-finite verb complement-taking verbs, like *to try/hate/forbid to do (something)*, cannot contain other verb forms. (Present participles are encoded as adjectives in all German corpora.) Auxiliaries and modals, which are included in LS, can dominate non-finite verb forms to build the present perfect tense and specify the modality of other verbs, respectively. Contrary to expectation (that finite verb forms would be most frequent in LST), the frequency of finite verb forms in LST is between the frequencies in VERBMOBIL and TüBa-D/Z. This finding can be attributed to the presence of a higher number of present perfect tense constructions in LST than in the non-LS written corpus (cf. the V.PP frequencies). The frequency of infinitives (V.INF and V.IZU, i.e., infinite with *zu* 'to', as in *Man hat so versucht*<sub>FIN</sub> *Corona aufzuhalten*<sub>IZU</sub> "One has tried to stop Corona in this manner") is similar in LST and

VERBMOBIL. However, the number of modals that are likely to entail a lexical verb is higher in LST (as a variety of written text) than in VERBMOBIL. As mentioned above, specific lexical verbs can dominate non-finite verbs with POS = V.IZU. The verb lemma *versuchen* 'to try' occurs 16 times in LST, 26 in VERBMOBIL, and 472 in TüBa-D/Z. The frequency in LST resembles that in VERBMOBIL, and the infinitive construction is similar to that of modals. We therefore add complement-taking verbs that belong to the restricted LS vocabulary to ELS. However, constructions with *um zu* 'for the purpose of/in order that' (KOU1), which occur 46 times in LST, 115 times in VERBMOBIL, and 2,426 times in TüBa-D/Z, are not added to ELS. The frequency of KOU1 is 50% lower in LST than in VERBMOBIL. Moreover, the construction can straightforwardly be segmented into: 'for the (following) purpose/thereby/!.. // and a main-clause construction, without obstructing the train of thought. For example, the sentence *Es braucht*<sub>FIN</sub> *Zeit um*<sub>KOU1</sub> *sich zu erholen*<sub>INF</sub> 'It takes time to recover' is divided into the following three lines: *Es braucht*<sub>FIN</sub> *Zeit. // Damit/Bis(:) // Sie erholen*<sub>FIN</sub> *sich (wieder).*

The preterite tense, passive voice, and subjunctive mood are forbidden in LS. The preterite occurs very infrequently in LST, and most of the 800 cases pertain to auxiliaries and modals. The few lexical verb cases can be replaced by present perfect tense forms (included in LS) without the meaning being changed. To support this argument, we searched VERBMOBIL for preterite forms of the three most frequent lexical verb lemmas according to Kempen and Harbusch (2019). For *sehen* 'to see' and *machen* 'to make,' no incidences were found. Forms matching *ging/-st/-t/-et/-en* of *gehen* 'to go' occur 31 times; however, only half of the matches are related to preterite forms. These all occur in the idiomatic phrase *das ging schnell/gut* 'that went quickly/well'. The other cases match subjunctive mood forms referring to potential time slots/connections/etc. As all verb forms of auxiliaries and modals appear with high frequency, and sentences are shortened by the use of the preterite (instead of present



perfect) tense, we include the preterite for VA and VM lemmas in ELS. The frequency of passive constructions is low in all three corpora. Given that such constructions are of medium difficulty according to the LeiSA study and that it is often hard to find a simple reformulation in the active voice that conveys the same nuance of meaning, we include passive constructions in ELS.

We noticed that nearly all subjunctive mood cases in LST are forms of auxiliaries or modals (e.g., *wären* ‘would be,’ and *möchten* ‘would like’). These are frequent word forms in German: there are 7,836 occurrences in VERBMOBIL for the rough search pattern POS = VM.\* or VA.\* with the word form matching the prefix = *möcht.\* / könnt.\* / würd.\* / wär.\* / hätte.\**. We therefore include the subjunctive forms of VA and VM in ELS. Constructions with double objects are of medium difficulty, according to the LeiSA study. Such constructions are equally rare in all three corpora examined in this study. However, common verbs like *geben* ‘to give (somebody something)’ qualify for this construction. As these verbs are included in LS, no extension is required.

**Phrase and sentence complexity.** A treebank search allows for very detailed syntactic specifications. However, it is necessary to keep in mind that the three inspected treebanks are differently encoded in this respect. For example, a noun phrase (NP), i.e., cat = NX in VERBMOBIL and TüBa-D/Z, covers constituents that differ from the nodes at the ends of edges labeled subj, obja, pn, etc., in PARZU. In an NX, adverbs can be seized. In PARZU, adverbs—if not in the frontfield—belong to the sentential level. For example, the phrase *auch schon viele barriere-freie Gebäude* ‘also already many barrier-free buildings’, occurring in the midfield of a clause, is assigned to three constituents on the sentence level in PARZU, whereas in the two other treebanks, the phrase is assigned to one NX node. There is no simple solution to this problem without manually inspecting all adverbs in LST. Hence, not entirely accurately, but in line with the characterization in the LeiSA study, we define phrase complexity by a simplistic dichotomy with respect to a length: We distinguish *short* (up to three words) from *long* (more than four words) constituents. This concept translates to sentence-complexity calculations: We define *short* sentences as containing up to nine words (assuming that these occur in no more than three constituents, each comprising no more than three words) and *long* sentences as containing more than nine words. (Note that punctuation is excluded from consideration here because any phrase-level search refers to an inner node of a corpus graph. In each of the three treebanks, punctuation is governed by the root node—cf. the example trees provided in the **Supplementary Material**. The discontinuous positions of punctuation symbols in the surface word order can be accessed by the list of word forms in the TIGER-XML format.) For example, *Bis zu diesem Gehalt zahlt man Beiträge* ‘Up to this income, you pay contributions’ is a short sentence, but *Die Krankenkasse oder die Agentur für Arbeit zahlt die Beiträge für sie*. ‘The health insurance or the employment agency pays the contributions for you.’ is a long one. On average, this simple distinction identifies complicated constructions: in our examples, a sentential modifier PP and an NP-modifying PP in an NP coordination, respectively.

In LS, no explicit length restriction for phrase complexity is stated. However, the LeiSA study qualifies as easy only those phrases with no more than three elements; therefore, no new constructions are added to ELS, and we omit detailed numbers here (see the **Supplementary Material** for the frequencies of the three argument NPs: subject [SB], indirect object [IO], and direct object [DO]—important elements for sentence understandability) and move on to investigations of sentence complexity.

We first quantify the frequency of the canonical SVO word order in main declarative clauses (see the upper panel of **Table 7**). We restrict the search to frontfields at the onset of a sentence to abstract away from elided constituents. (*Forward Conjunction Reduction* and/or *Gapping* are the only ellipsis phenomena that can elide the left periphery and only in the second conjunct of a coordinated sentence; see Ross, 1967). Moreover, according to Temperley (2019), the most complex constructions occur at the onset of a sentence because more mental capacity is available here.

In line with LS, the LeiSA study allows only SVO word order; even mild forms of topicalization were judged to be very difficult for LS readers. Unexpectedly, the frequency of the canonical word order is found to be very similar in all three treebanks; constructions with the SVO word order comprise only half of all constructions in the LS corpus, i.e., the other half are very complicated for the target readers. Clearly, the standard German writers of LS texts adhere to the standard rules of German discourse structure. In mild cases, a one-word constituent occupies the frontfield (cf. example (3) above). However, we also found complex frontfield fillers, such as conditional clauses, sentential subjects, and objects. For example, in a sentence (9), the fronted object is interpreted as the argument of the finite verb form *wollen* until *barriere-frei* occurs (i.e., this is a garden-path sentence). Given the difficulties arising from simple deviations from the canonical word order, we do not report the frequencies of individual constructions here, and no deviations are included in ELS.

- (9) Das<sub>ACC</sub> wollen<sub>FIN</sub> wir in NRW barriere-frei machen<sub>INF</sub>.  
This want we in NRW barrier-free make  
‘These things we want to make barrier-free in NRW.’

For the canonical word ordering, the average length of the frontfield is longer in LST and TüBa-D/Z, the two written varieties—as expected. The subjects in VERBMOBIL are extremely short due to the use of personal pronouns in dialogue. (Note that VERBMOBIL often has discourse markers, self-repairs, etc., at the onset of a sentence. Therefore, the total numbers for LST and VERBMOBIL diverge more here than in other tables.)

For subordinate clauses, we distinguish between relative clauses and subordinate clauses starting with a subordinating conjunction (KOUS) in the lower panels of **Table 7**. Both constructions are forbidden in LS. According to the LeiSA study, subject-relative clauses are of low difficulty for LS readers, whereas any other type of relative clause (i.e., a relativizer in the dative or accusative case due to its grammatical function or dominating preposition, respectively) is extremely difficult. Surprisingly, both types occur with approximately equal frequency in LST and TüBa-D/Z. As

expected, the relative clauses in LST are considerably shorter than those in TüBa-D/Z. Given their frequency and the LeiSA evaluation, we opt to include subject-relative clauses in ELS (although this adds a VF construction to the included word order patterns); however, we suggest that such clauses are not discontinuous and that they should be short.

All types of subordinate clauses are forbidden in LS. In the LeiSA study, subordination with the temporal conjunctions ‘while/after’ is considered of low difficulty. Although ‘while’ does not occur in LST, *nachdem* ‘after’ is used five times. The conditional conjunction *wenn/falls* ‘if’ is by far the most frequent (see **Table 7**). Other conjunctions used in LST (with their respective frequencies) include: *als* ‘when’ (8); *bevor* ‘before’ (18); *bis* ‘until’ (37); *damit* ‘so that’ (5); *indem* ‘by’ (2); *nachdem* ‘after’ (5); *ob/obwohl* ‘whether’ (204); *seit* ‘since’ (1); *solange* ‘as long as’ (3); and *weil* ‘because’ (180).

Rather than including specific subordinate clause types, we suggest adding all subordinating conjunctions to ELS. However, the conjunction and the sentence should be presented in two separate consecutive lines, and the trailing sentence should have main clause word order (cf. example (4) for the paratactic conjunction *denn*, which always entails main clause word order). The same construction works with the synonymous subordinating conjunction *weil*. In VERBMOBIL, the subordinating conjunction *weil* ‘because’ occurs in half of the cases with SVO order (Kempen and Harbusch, 2016). (This phenomenon is widely studied as the *weil*-V2 phenomenon in spoken standard German; see Reis (2013) for a broad overview.) This strategy also covers the highly difficult construction of dependent *that* clauses, which occur in LST slightly less often than in VERBMOBIL. This construction can straightforwardly be avoided by replacing *that* by a colon. The content of the *that* clause is presented as a main clause in the canonical SVO order. Thus, we do not add this construction to ELS.

**Coordination and ellipsis.** Coordination and ellipsis are of medium difficulty, according to the LeiSA study. However, the tested examples are very simple. Coordinations in LST consist of very long lists, as illustrated above in example (8). Often, formal definitions are replaced by long lists of examples, probably to avoid the use of overcomplicated technical terms. Therefore, although we do not add any new constructions, we recommend using only short lists of coordinated constituents.

The same holds for ellipsis. The use of ellipsis in spoken and written text (see corpus studies into VERBMOBIL (Harbusch and Kempen, 2009) and the TIGER treebank, another syntactically annotated German newspaper corpus (Harbusch and Kempen, 2007)) goes beyond the scope of very limited Forward Conjunction Reduction restricted to the subject, which prevails in LST, and which was the only type of ellipsis evaluated in the LeiSA study. As this construction is judged to be of medium difficulty, and it can be circumvented by explicitly repeating or pronominalizing the subject, we choose not to add it to ELS.

### 2.3.3 Summary of the Constructions in Extended Leichte Sprache

**Table 8** sums up all extensions included in ELS that we proposed in the previous section. Now that the range of syntactic constructions of

**TABLE 8 |** Summary of constructions in Extended Leichte Sprache that extend Leichte Sprache.

Phenomenon	Usage
Negation	Restricted to <i>nicht</i> as verb modifier and <i>kein</i> <sub>inflected</sub> as determiner or substituting pronoun
POS = APPRART	Agglutination of the definite article following the preposition
POS = PRO	NP filled with a substituting pronoun
Preterite tense	Restricted to auxiliaries and modals
Subjunctive mood	Restricted to auxiliaries and modals
Passive voice	Restricted to forms of <i>werden</i>
Complement-taking verb	Restricted to verb lemmas in the vocabulary of LS that dominate a nonfinite verb form with the word ordering SV <sub>finite</sub> OV <sub>nonfinite</sub>
Subordination	<i>dass</i> ‘that’ is replaced by a colon; any other conjunction obtains a separate line; the constituents of the subordinate clause are presented with V2-word order in a new line
Relative clause	Restricted to subject-relative clauses, however, not discontinuous

ELS has been defined, we detail how our system supports the writing of ELS text.

## 3 EXTENDEDEASYTALK

First, we present the state of the art in writing support tools. In **Section 3.2**, we give an intuitive impression of how *ExtendedEasyTalk* works. **Sections 3.3** and **3.4** go into the details of the computational linguistic mechanisms used to support the writing of a sentence and the production of a sentence-coherence element, respectively. In **Section 3.5**, we sketch the active mode of text-production teaching. Finally, we present the results of our evaluation with different user groups.

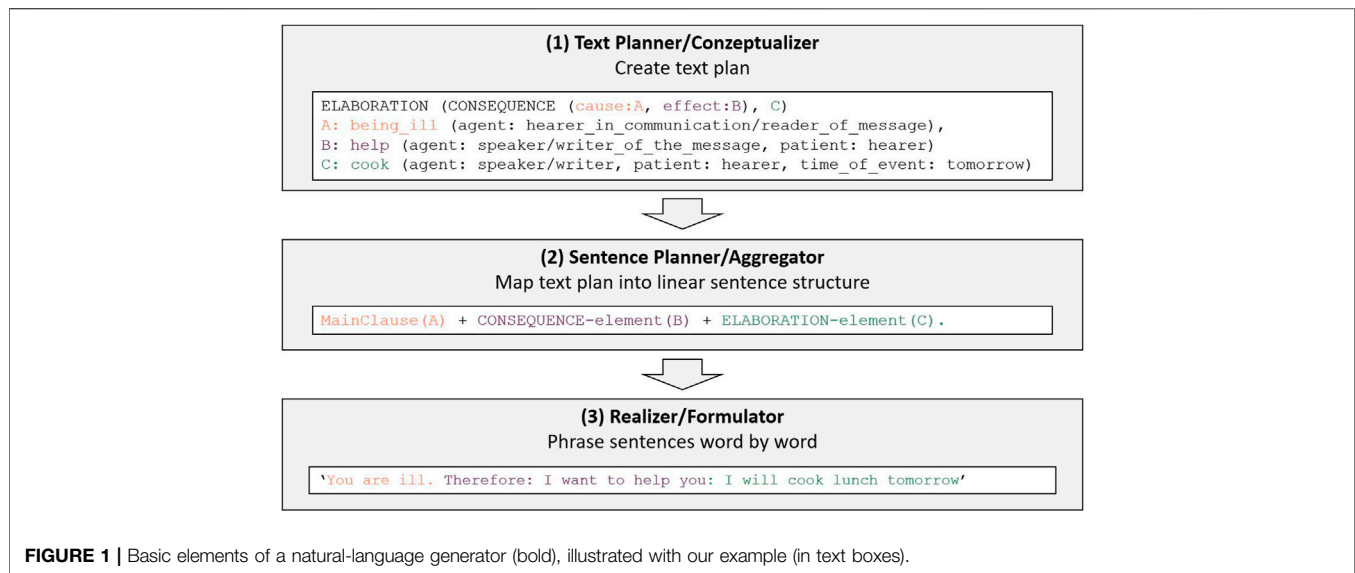
### 3.1 The State of the Art in Writing Support Tools

First, we describe the state of the art in technical writing support in the overall area of AAC. Then, we focus on NLG-based approaches. This latter section provides the blueprint for a knowledge-based automatic natural-language generator, enabling us to refer back to concepts used in our system.

#### 3.1.1 Augmentative and Alternative Communication

For people with congenital or acquired communication impairments, the use of *Augmentative and Alternative Communication*<sup>3</sup> (AAC) is often an essential means of inclusion, i.e., for self-determined participation and self-expression. AAC offers a wide range of communication techniques, including gestures,

<sup>3</sup><https://isaac-online.org> (information provided by the *International Society for Augmentative and Alternative Communication*).



signs, and graphic symbols, as well as technical communication aids. Technical AAC solutions range from simple concatenations of symbols for needs-based, functional communication using limited vocabulary to complex customizable systems (see Lancioni et al. (2019) for a detailed survey).

Here, we focus on AAC systems that aid users with at least low literacy skills both to express basic necessities and to write about topics that create social closeness, for example, to share personal information and experiences (see, e.g., Light et al., 2019). The rich morphology and the relatively complex word-order rules of German complicate the generation of useful and grammatically correct suggestions. Commercial systems that go beyond functional communication include MindExpress, Gateway, and Snap Core First<sup>4</sup>. These systems essentially concatenate words, word groups, and symbols into sentences, thereby providing basic linguistic support, such as adaptive word prediction and automatic inflection for simple constituents. Technical AAC solutions are currently evolving rapidly and are increasingly available on mainstream devices (e.g., smartphones and tablets; Light and McNaughton (2012)). All popular free apps for German allow users to access large customizable vocabularies of (visual) symbols. However, they do not provide well-founded linguistic support for sentence construction and/or text production (cf. LetMeTalk and SymboTalk<sup>5</sup>). Importantly, these systems are mainly intended for direct (face-to-face) communication.

### 3.1.2 Natural Language Generation

There is an increasing demand for language support through linguistic processing by computer. However, the currently

available AAC systems do not exploit the full potential of computerized linguistic processing (Waller, 2019). The technical authoring support available for LS includes tools for automatic text simplification based on parsing (for German see, e.g., Suter et al., 2016) and text validation tools (see, e.g., LanguageTool<sup>6</sup>, a system that flags violations of the LS rules).

To the best of our knowledge, there is no recent NLG-based text-production system customized to AAC-user needs (cf. the pioneering approach by Demasco and McCoy (1992); Gatt and Krahmer (2018), who illustrate the potential of NLG systems in general; and G2.com (2021), which subcategorizes systems as “Highest rated/Easiest to use/Free” and provides links to writing support based on NLG).

In the following, we outline the steps from a speaker’s intention to a context-sensitive utterance to allow us to highlight the NLG concepts we employ in *ExtendedEasyTalk*. **Figure 1** shows the typical three-stage pipeline architecture of a declarative text generation system (see, e.g., Figure 3 on page 13 by Reiter and Dale, 2000), illustrated in terms of example (1-a) above. The overall input to the Text Planner/Conceptualizer (component 1) is encoded as the speaker’s *goal*:

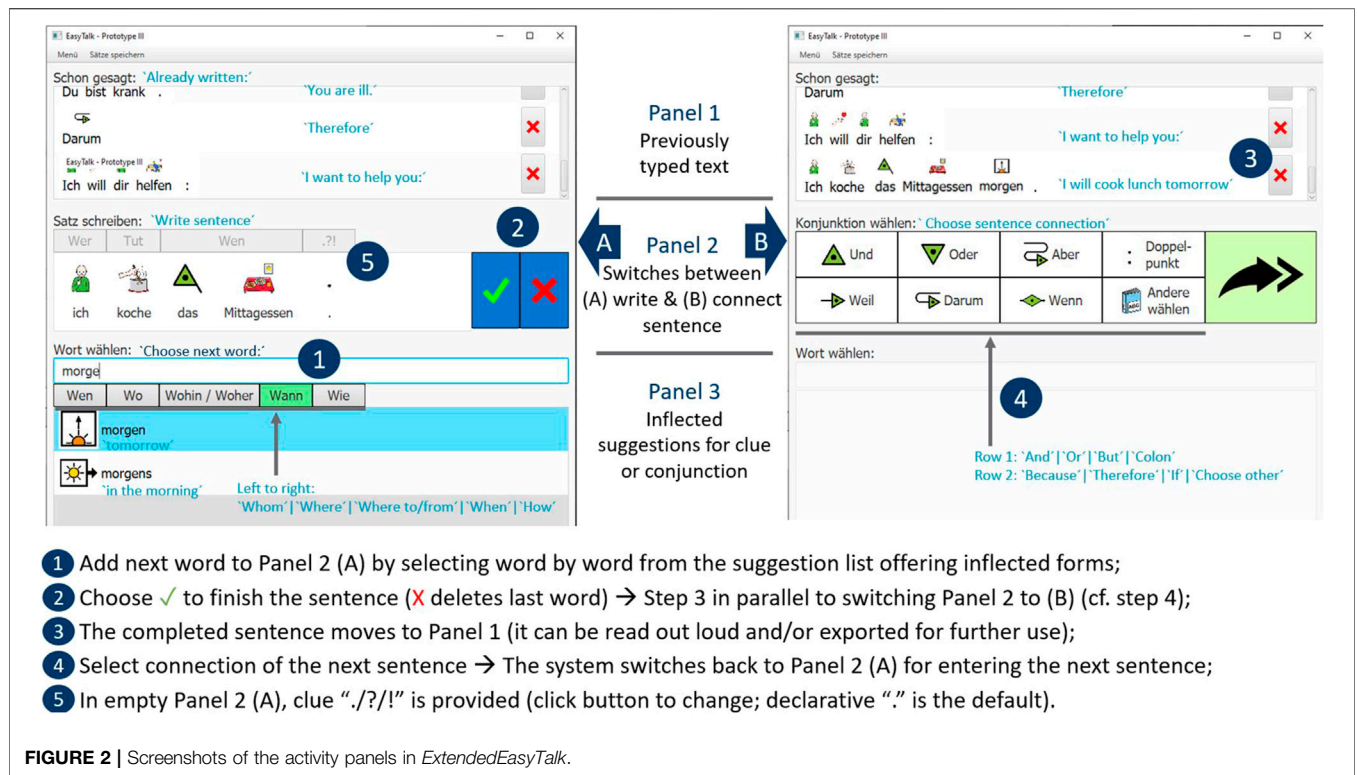
INFORM(*S*, *H*, KNOW(*H*, cooking(agent\_of\_action\_cooking: *S*, object: lunch, time: tomorrow))).

*S* refers to the speaker, and *H* to the hearer. The goal is an INFORM-speech act, i.e., *S* wants *H* to know something that *S* assumes *H* does not yet know. (In a REQUEST-speech act, *S* wants to obtain new information that *S* assumes *H* knows, often resulting in a question like “*Can you please tell me what you would like to eat for lunch tomorrow?*”) Roughly speaking, the goal highlights the discrepancy between the knowledge bases of *S* and

<sup>4</sup>www.jabbla.com (MindExpress); www.gatewaytolanguageandlearning.com (Gateway); www.tobiidynavox.com/pages/snap-core-first (Snap Core First).

<sup>5</sup>www.letmetalk.info (LetMeTalk); www.symbotalk.com (SymboTalk).

<sup>6</sup>www.languagetool.org/de/leichte-sprache (LanguageTool).



*H*, which should be removed through a series of speech acts, i.e., the not-yet-verbalized conceptual messages (propositions dealt with the Conceptualizer). Ideally, after the delivery of the message by *S*, *H* knows all communicated facts (and facts that can be inferred by *H*). This task requires separate representations of the speaker’s and the hearer’s knowledge about the current situation and about their presupposed *world knowledge*, respectively (including the implications of/inferences from all the facts).

In the example, we assume that *S* notices that *H* is ill. *S* wants to indicate to *H* that *S* is aware of this fact. By explicitly informing *H* that *S* knows that *H* is ill, the speaker enriches the utterance with a known fact to make the context/intention of the utterance clear, thereby creating an overall *discourse structure*. Because of the close personal relationship between *S* and *H*, *S* decides to help by preparing lunch for *H* the next day. This plan results from the world-knowledge fact that relieving an ill person of a task helps that person to rest and recover.

The Conceptualizer decides which information should be communicated. In our example, the facts *A*, *B*, and *C* are selected (the propositions are rendered here in the form of sentences, abstracting away from the logic representation, and detailed argumentation; instead, each proposition is supplemented by the intended interpretation of the hearer):

*A: Du bist krank* ‘You are ill’: mutual agreement about the context of the utterance

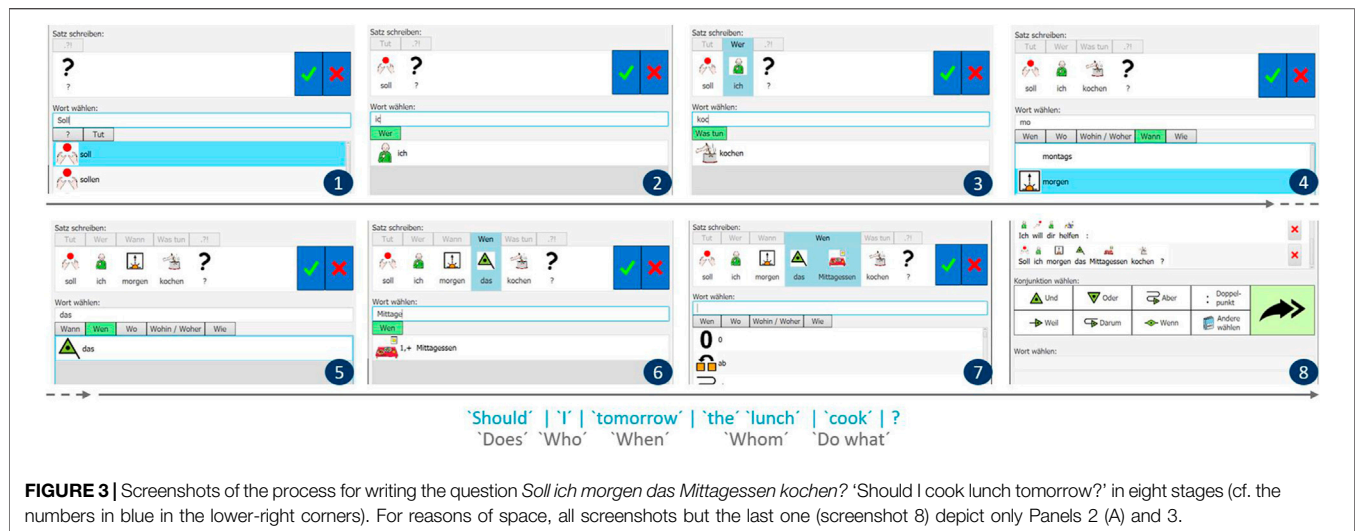
*B: Ich will Dir helfen* ‘I want to help you’: reason for a proposed action

*C: Ich koche morgen das Mittagessen* ‘I will cook lunch tomorrow’: communication of the planned action

Importantly, propositions do not stand in isolation, but in the relationship, in order to express the discourse structure/the speaker’s intention. A widely used technique for this purpose is that of *Rhetorical Structure Theory* (RST; see Hovy, 1988 and Mann and Thompson, 1988). Two important examples of relations between propositions are ELABORATION and CONSEQUENCE. The resulting hierarchical structure of interrelated propositions is called the *text plan*. Text plans are handed over to the Sentence Planner/Aggregator module (component 2), which has the task of linearizing the hierarchical structure. The linearization process involves, among other things, the insertion of coordinating and subordinating conjunctions and other lexical items that instantiate RST relations (although not all RST relations need to surface explicitly in the final text). In terms of our example:

- *A* is realized as the main clause;
- CONSEQUENCE(*B*) is realized by the causative adverb *therefore* preceding the main-clause realization of *B*, and
- ELABORATION(*C*) is realized by a colon preceding the main-clause realization of *C*.

Component 3, the Realizer/Formulator, provides the subsequently generated text—one of the many realization options of the overall generation system.



In the NLG nomenclature, all AAC systems mentioned in Section 3.1.1 are restricted to formulator problems. In the following, we illustrate the potential of using concepts from all three generation steps in our system.

### 3.2 A Writing Session With ExtendedEasyTalk

Let us familiarize ourselves with the assisted text-production process of *ExtendedEasyTalk* through an outline of its five essential steps (cf. the numbers in blue circles in Figure 2).

The system permanently displays three panels. Panel 1, at the top, contains the previously written text. In the middle panel, Panel 2 (A) accumulates the word forms of a sentence chosen word by word from Panel 3, and Panel 2 (B) provides predefined sentence connectors. Panel 3 offers the list of suggestions for the currently typed string. The example depicts the fact that the user has already produced a sequence of sentences: *Du bist krank. // Deshalb // Ich will dir helfen*: 'You are ill. // Therefore // I want to help you.' in the left-hand side of Panel 1. The sentence under construction, *Ich koche Mittagessen*. 'I will cook the lunch', is displayed in Panel 2 (A). The user has two options for how to proceed: the sentence is either continued or finished.

In step 1, a new word is selected to be added to the sentence using Panel 3 (cf. lower-left corner). *ExtendedEasyTalk* offers inflected word forms according to the wh-cue header in green (here, **Wann** 'when' refers to the grammatical function provided in simplistic wording) matching any (possibly empty) input string the user types in Panel 3. In the figure, the user has typed *morge* 'tomorrow'. As predictions, the system presents inflected completions from the lexicon, according to the grammatical function referred to by the active cue word. In the example, two alternatives are retrieved. The top-most element (shown in blue) has been selected here. Consequently, the element moves to Panel 2 (A) (the result is not depicted in the static figure). In step 2, the user finishes the process of sentence production by selecting the green checkmark in Panel 2 (A). In response, the completed sentence moves to Panel 1. The result is depicted on the right-hand side of

Panel 1 (marked as step 3), where the newly completed sentence *Ich koche das Mittagessen morgen*. 'I will cook lunch tomorrow.' is appended to the previously written text. The user can scroll through Panel 1 to look back within the flow of thoughts. A read-aloud function serves to remedy reading deficits. To support writers with low literacy skills, AAC symbols<sup>7</sup> supplement each word form. The produced text can be exported from the system for further use, with or without symbols. When the completed sentence is added to Panel 1 in step 3, Panel 2 simultaneously switches to the menu (B), offering sentence connectors (step 4). After an element that meets the user's communicative intention is chosen (or skipped *via* the large green arrow button), it is also moved to Panel 1; the system then switches Panel 2 back to (A) for the next sentence to be entered (step 5). At this point, only the punctuation cue is provided in order for the sentence type of the next sentence to be selected (cf. Figure 3 for an illustration of how a question is typed).

In the following two sections, we elaborate on the computer-based linguistic support in the two phases of the text production process, i.e., within a sentence and between consecutive sentences. A particular highlight of the descriptions comprises arguments for an adequate user interface (UI) that, in addition to supporting low literacy skills, must compensate for factors such as working memory deficits within the target user group of people with a wide spectrum of cognitive impairments and/or learning difficulties.

### 3.3 Fast and Correct Extended Leichte Sprache Sentence Production

With respect to the goal of fast and correct typing, using Panels 2 (A) and 3 (cf. the left-hand side of Figure 2), we employ a variant of an NLG formulator. Its goal is to build up a derivation tree

<sup>7</sup>Here, we use the ARASAAC symbol set: [www.arasaac.org](http://www.arasaac.org). The symbol set can be adapted to the preferences of the user. Moreover, advanced users can switch this mode off.

based on the rules of a syntactic grammar (here, ELS constructions; if desired, the declarative grammar can easily be restricted to pure LS rules) so that syntactic correctness is automatically maintained. Based on this representation, the system produces correctly inflected word forms.

In NLG, the formulator usually administers only the best sentence representation; there is no UI enabling the selection of another option (*paraphrase*). We adopt a slightly more flexible formulator approach, developed for L2 learners of German. COMPASS (Harbusch et al., 2007, 2014) is a *natural-language paraphrase generator* that constructs the sentence the user has in mind in a step-by-step dialogue. The process is called *scaffolded writing*, in reference to the fact that the system is able to maintain the syntactic correctness of the construction the user is typing after the user has specified its grammatical function (Harbusch and Kempen, 2011). COMPASS is based on the grammar rules in *Performance Grammar*, a psycholinguistically motivated grammar formalism (Harbusch and Kempen, 2002; Kempen and Harbusch, 2002). The separation into distinct dominance and word order rules enable a flexible sentence-production process to suit the user's preferences. For example, the user can enter all arguments first to empty the short-term memory and then fully concentrate on arranging the constituents according to the intended discourse structure. Revisions made throughout the sentence at any point in time are retained. Upon request, the system reports whether or not a given construction is authorized by the grammar.

In essence, the overall lexicon covers the German CELEX (Gulikers et al., 1995). To obtain a reasonable suggestion list, this lexicon is restricted to L2-learner level A1/A2 in *ExtendedEasyTalk*. The system can also be adapted to the user's personal vocabulary (e.g., to include proper names of protagonists or places) or specific contexts (e.g., for school purposes).

The set of declarative rules applied by *ExtendedEasyTalk* is restricted to ELS constructions (cf. Table 8). For example, the range of verb forms is restricted to the active voice, indicative mood, and present and present perfect tenses. For auxiliaries and modals only, the preterite and the subjunctive mood are also offered. The system favors non-inversion word order. See Table 9 for the order in which the constituents are presented in main declarative sentences. To provide an intuitive UI, it is crucial to avoid linguistic terms; therefore, we use cues in the forms of interrogative pronouns, as outlined in Column 1, to communicate with the user about grammatical functions and maintain scaffolded writing. (This technique resembles elementary school exercises for identifying grammatical function fillers in a sentence.) In return, the system is enabled to propose correctly inflected forms.

For the cues in Panel 2 (A), we illustrate how much information the system controls for word forms rather than showing the overall derivation tree here:

**Wer** 'who<sub>nom</sub>' = **Subject**: *Ich*PPER:1stPers,Sing,Nom

**Tut** 'does' = **Verb**: *koche*VVFIN:1stPers,Sing,Pres,Ind,ActiveVoice

**Wen** 'whom<sub>acc</sub>' = **Direct object**: *das*ART:Def,Neut,Sing,Acc *Mittagessen*NN:Neut,Sing,Acc

Now, we illustrate how this information is collected in an easy and intuitive step-by-step manner. Initially, the system presents the

**TABLE 9** | List of constituents in a main declarative sentence (in the top panel, subject and finite verb forms are obligatory; the second panel enumerates further arguments/valency-frame fillers of the finite verb; and in the lower panel, adjuncts/modifiers are enumerated). Column 1 provides the cue words to be displayed as headers in Panels 2 (A) and 3.

Cue	Automatically inflected filler
<b>Wer</b> 'who <sub>nom</sub> '	Elements of the <b>SUBJ</b> ect in nominative case
<b>Tut</b> 'does'	<b>FIN</b> ite verb form in active voice, present tense, coinciding in person and number with the subject
<b>Wem</b> 'whom <sub>dat</sub> '	Elements of the Indirect <b>OBJ</b> ect in dative case
<b>Wen</b> 'whom <sub>acc</sub> '	Elements of the <b>DIR</b> ect <b>OBJ</b> ect in accusative case
<b>P<sub>r</sub> was</b> 'P <sub>r</sub> what'	Elements of the <b>PRE</b> positional <b>OBJ</b> ect in the case <b>P<sub>r</sub></b> , the instantiated preposition requires
<b>Was tun</b>	<b>P</b> ast <b>P</b> articiple in case the finite verb form is an auxiliary or
'what to do'	<b>INF</b> initive in case the finite form is a modal or <b>INF</b> initive_ with_ <b>ZU</b> in case the finite form is a complement-taking verb
<b>Wann</b> 'when'	Elements of <b>MOD</b> ifier_time
<b>Wo</b> 'where <sub>loc</sub> '	Elements of <b>MOD</b> ifier_location
<b>Woher/-hin</b>	Elements of <b>MOD</b> ifier_direction from/to
'where <sub>dir</sub> '	
<b>Wie</b> 'with what'	Elements of <b>MOD</b> ifier_instrument

cues for all sentence components according to the canonical word order of the chosen sentence type. In a declarative main clause presupposing SVO word order, the subject is entered first. Based on the subject's number and person features, the system provides only correctly inflected verb forms for any typed word prefix managed by the cue **Tut** 'does' in Panel 3. In the example, the first two cues are filled with *ich koche* 'I cook.' In the list of choices, all forms with a separable verb prefix (SVP) (e.g., *koche ab* 'to boil off') that are covered by the currently selected lexicon are presented to the user. In the example, the verb lemma *kochen* (without any SVP) is chosen. If a verb with an SVP is selected, *ExtendedEasyTalk* assigns its word order position automatically (cf. *ich koche das Wasser ab* 'I boil off the water'). If the finite verb is an auxiliary, modal, or complement-taking verb included in ELS (e.g., *to try*), the sentence can continue with either a direct object or another verb with its own valency frame to be filled (e.g., *ich will ein Eis* 'I want an ice cream' vs. *ich will Ball spielen gehen* 'I want to go play ball'). This decision is presented to the user in a simple manner as a choice between the cues **Wen** 'whom<sub>acc</sub>' and **Tut was** 'does what', respectively (see Steinmetz and Harbusch (2020) for details of how the user is supported in filling the valency frame of recursively added verbs).

After the verb is entered, the system keeps track of the overall valency restrictions/arguments provided in the lexicalized grammar, and every word form is supplemented with the appropriate syntactic structure. Incompletely filled valency frames cannot be moved to Panel 1; i.e., only correct sentences can be typed in *ExtendedEasyTalk*. In the example, a direct object cued by **Wen** 'whom<sub>acc</sub>' has been filled with *das Mittagessen* 'the lunch', i.e., the cue **Wen** overarches the whole direct object. In Panel 3, the grammatical function currently active in Panel 2 (A) can be expanded (i.e., so that the same cue is active in Panels 2 (A) and 3) until the user selects another cue in Panel 3 or finishes the sentence. Modifier/adjunct cues are facultative. Displaying them should prompt the user to

supplement the sentences properly with audience-design information, such as the time and place of an event (cf. **Section 3.5**). In tests with L2 learners, beginners completely ignored this offer without feeling disturbed.

In our example, by selecting the modifier cue **Wann** ‘when’ (displayed in green), the user decided to add *tomorrow* as a temporal specification that the reader should know. At the point shown, the user has typed “*morge*” in the text-input field of Panel 3. Accordingly, a choice list presenting only temporal expressions is retrieved from the lexicon matching the current input string. In the figure, two items differing in inflected endings qualify as matches for “*morge*”. The user navigates the completion list by scrolling to the intended form.

The typing speeds of all users, not only LS writers, are supported by prediction/completion lists (cf. typing on reduced keyboards on cell phones). The structure of Panel 3 borrows this concept. For any string prefix—even an empty one<sup>8</sup>—a suggestion list is displayed according to the active cue. *ExtendedEasyTalk*’s inflected suggestions speed up typing by unifying the two-stage process of selection and manual morphological adaptation. Hence, not only is syntactic correctness maintained but also typographical errors are also avoided. The weakness of this method became apparent to us during the evaluation of the system (cf. **Section 3.6.2**). Spelling deficits lead to empty lists and lengthy trial-and-error attempts. This issue will therefore have high priority in our future work.

In order to sum up the supportive features of *ExtendedEasyTalk*, we describe the typing of the six-word question *Soll ich morgen das Mittagessen kochen?* ‘Should I cook lunch tomorrow?’ in eight steps in **Figure 3**. The punctuation cue in Panel 2 (A) provides a declarative main clause by default. The user can switch to any other sentence type by *scanning*, i.e., repeatedly pressing the punctuation cue button until the correct choice appears. The word order, i.e., the order of the cue words in Panel 2 (A), is adapted according to the selected sentence type. The chosen punctuation mark automatically remains sentence-final at the end of the sentence during the process of typing the sentence.

In the example, the user has selected a yes/no question. For the typed string prefix ‘soll’, the system ranks the forms *soll*<sub>1st/3rdPers,Sing</sub> and *sollen*<sub>1st/3rdPers,Plur</sub> ‘should’ in the topmost positions. When the verb is typed in a sentence-initial element, i.e., lacking the features of the subject, the system cannot do any better. Thus, all possible (ELS-approved) verb forms have to be enumerated. Next, the user is required to fill in the obligatory subject. Now, subject–verb agreement can be used to filter the subject forms according to the chosen inflected verb form *soll*<sub>1st/3rdPers,Sing</sub>. (cf. step 2). In line with the typed string prefix ‘ic,’ the personal pronoun *ich* ‘I’ is the only option in the completion list. In step 3, the user has to follow the obligatory cue **Tut was** ‘does what’ elicited by the modal finite verb form. (The lexicon used here does not contain any lemma where *kochen* holds a separable verb prefix.) In step 4, the cues in Panel 3

present a list of the next constituents. The user can omit the facultative indirect object (**Wem**). In our example, the user decides to add the time of the event, *morgen* ‘tomorrow,’ before the direct object, *das Mittagessen* ‘the lunch.’ Note that advanced users can deviate from the default order by jumping directly to a certain cue in the list; the correct overall German word order is maintained by the system regardless. From step 5 onward, the filled cue **Wann** ‘when’ is displayed in Panel 2 (A). In steps 5 and 6, the user enters the direct object. In step 7, the user operates the checkmark button, and the sentence from Panel 2 (A) moves to Panel 1. In parallel, Panel 2 (B) appears (this is discussed in more detail in the next section).

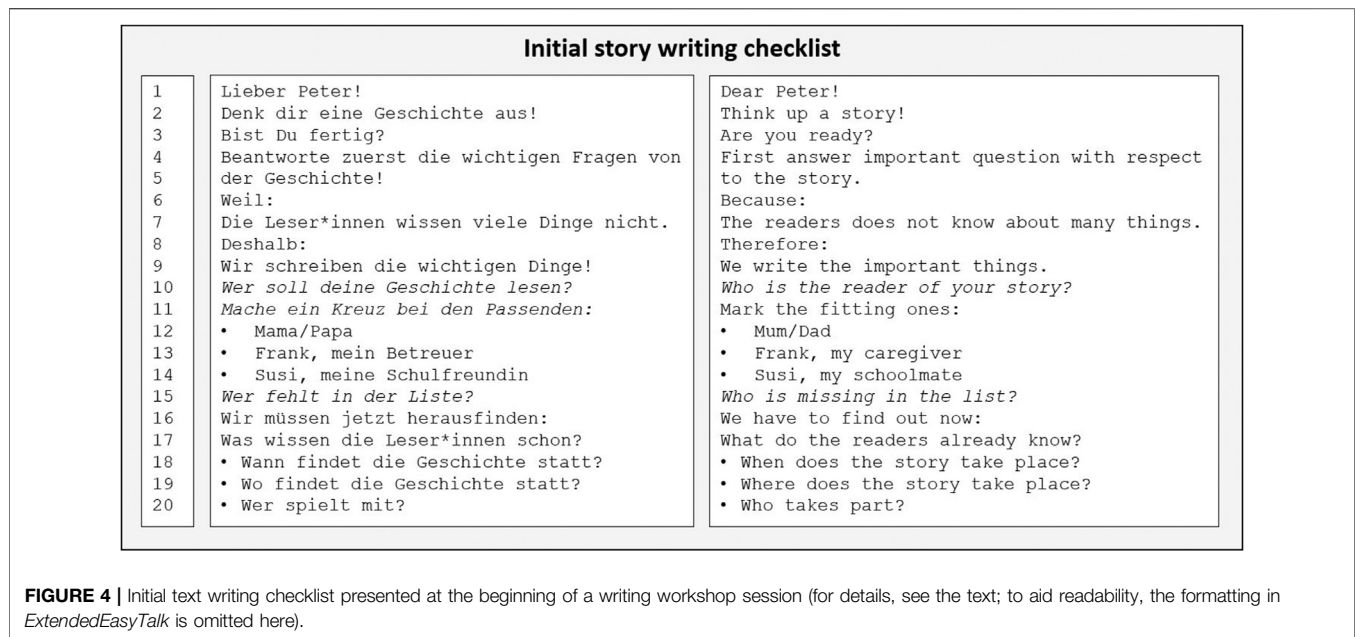
### 3.4 Production of Elements for Sentence Coherence

Writing support is not restricted to intra-sentential items. Text consisting of a series of simple sentences with canonical SVO order lacks flow, and the writer’s thoughts are only partially communicated. As in the conceptualizer of an NLG system, RST-like cues relating to the individual sentences should verbalize the user’s communicative goal. As mentioned in **Section 1**, techniques for exemplifying RST relations are learned in exercises for complex clause construction in school. However, complex clauses with informative conjunctions are not available in either LS or ELS.

Having noticed in our corpus study that constructions of the form (conjunction/adverb [possibly followed by a colon] // main clause) improve text understandability, we decided to add this concept to *ExtendedEasyTalk*. We assume that our users are familiar with the use/meaning of most conjunctions in German (cf. the frequencies of subordinating conjunctions, i.e., KOUS, in LST, provided in **Section 2.3.2**). Moreover, the LeiSA study evaluated subordinate clauses with *während/nachdem* ‘while/after’ as easy for LS readers, even with subordinate VF word order. Thus, instead of using technical terms like ELABORATION to refer to RST relations, we ask the user to select an appropriate conjunction/adverb. Whenever the user finishes a sentence (by pressing the green checkmark button; cf. step 2 in **Figure 2** and step 8 in **Figure 3**), Panel 2 switches to menu (B). This menu consists of nine buttons (pressing the green arrow button on the right side of the menu omits the addition of a connector). In accordance with suggestions made by AAC experts (cf. **Section 3.6.1**), we restrict the choice to those forms widely used under LS rules Netzwerk Leichte Sprache to avoid overtaxing the user.

We group the elements in the menu according to conjunction type. In the upper row, the coordinating conjunctions *und* ‘and’ *oder* ‘or’ and *aber* ‘but’, and the colon are provided. We realize that the colon is highly ambiguous in LS texts; however, its use is widespread (Bredel and Maaß, 2016, p. 254). We therefore offer this choice to prevent users from having to search for this option. In the second row, the user is presented with the subordinating conjunctions *weil* ‘because’ and *wenn* ‘if’, the adverb *darum* ‘therefore’, and a button *Andere wählen* ‘Choose other’. In our corpus study into LST, we observed further variation for POS = KOUS; hence, more advanced users can

<sup>8</sup>Cf. the active completion list of an empty word prefix in screenshot 7 in **Figure 3**. The list reflects the currently very limited context.



**FIGURE 4 |** Initial text writing checklist presented at the beginning of a writing workshop session (for details, see the text; to aid readability, the formatting in *ExtendedEasyTalk* is omitted here).

browse through all conjunctions. For consistency and overall ease of use of the system, Panel 3 provides a list of conjunction choices with the same selection options as for word forms in sentence typing. In case the option selected—either by button or in Panel 3—is a word, it is added as a separate line at the end of Panel 1 (cf. *darum* ‘therefore’ in the second line on the left-hand side of Panel 1 in **Figure 2**); the colon is appended to the last sentence in Panel 1 and replaces the previously written punctuation symbol.

So far, we have illustrated how users can use *ExtendedEasyTalk* to freely type ELS constructions. In the next section, we focus on the teaching of text production concepts by wrapping an active control structure around the key components for typing.

### 3.5 Control Mechanisms of the Writing-Workshop Mode in *ExtendedEasyTalk*

Prolific writers know that coherent, understandable text has to emphasize audience/reader-design concepts. *ExtendedEasyTalk* can teach basic writing workshop concepts; to do this, the system takes the initiative by asking questions (stated in ELS) at different stages of text production. For convenience, this mode can be easily ended or reactivated at any point in time.

At the beginning of a text, the user has to answer questions from a checklist (cf. **Figure 4** for an excerpt; the questionnaire presented to the user can be adapted to specific text genres). Depending on the user’s reading fluency, the questions can be read aloud to them (e.g., by a caregiver), or the read-aloud function of the system can be used to speed up the dialogue. As far as possible, the individual questions of the checklist offer a range of alternatives. Where this is not possible, the user types the answer using *ExtendedEasyTalk*. The answer lists can be pre-adapted to the current user, e.g., the names of the user (in our

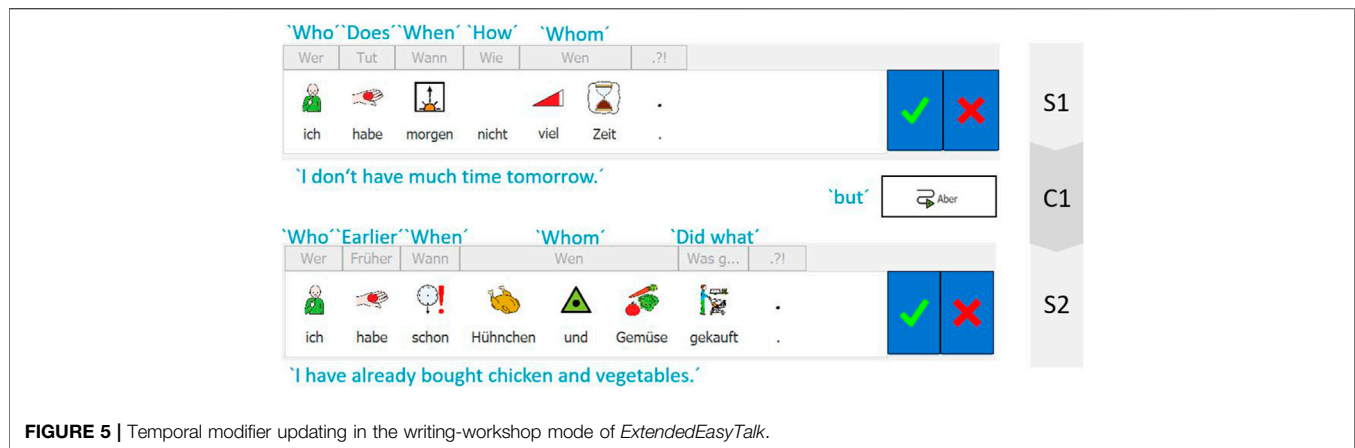
example, *Peter*, a male user) and caregivers, friends, teachers, etc. The dialogue starts with an introductory text (cf. lines 1–9). Lines 10–15 collect background information on the hearer/reader in an intuitive manner. Lines 16–20 show part of the collection of background information for the text the user would like to write.

To characterize all the protagonists in the list of actors (line 20) so that the reader can identify them clearly, a sequence of questions is asked. Different options are tested. Does the reader already know the name of the actor(s)? Can they be introduced by name? Can a characterization of the person(s) be added to enable the reader to become familiar with them (e.g., *Frank is my caregiver; Susi is my schoolmate*)? Such a session avoids the need for relative clauses (although these are included in ELS).

Similarly, the background of every sentence is established through questions referring to the adjuncts (cf. the modifier cues in **Table 9**). In the writing-support mode, the system will infer that the place and time have changed. Instead of simply displaying the temporal modifier cue (cf. step 4 in **Figure 3**), it will ask the user an explicit question to keep track of all changes or details unknown to the reader.

This looks cumbersome; however, the effort pays off when protagonists are referred to during sentence production. Not only can suggestions of personal pronouns be made by the system but also, in addition, the temporal and spatial modifiers are prefilled with the initial/most recent filler, ruling out wrong assumptions by the reader. In **Figure 5**, the first sentence, S1, refers to *tomorrow* due to the intervention of the system to initialize time and place. Without active intervention, the second sentence, S2, would also display the **Wann** cue filled with *tomorrow*. This leads the user to notice the clash with their intended content; in the example, the cue is revised to *already*. The





**FIGURE 5** | Temporal modifier updating in the writing-workshop mode of *ExtendedEasyTalk*.

user types S2 in the present perfect tense. The question of whether or not the system should actively provide tense suggestions remains open. We hesitate to make our system overly adaptive. Many users—irrespective of their specific user group—do not appreciate non-static UIs (Lee and Yoon, 2004).

Finally, we report our evaluation results.

### 3.6 Evaluation

As testing software with people with disabilities presents special challenges and organizational overheads (see, e.g., Henry, 2007 or Lazar et al., 2017), we decided to test as many facets of our system as possible with substitute users. For our target group, the initial impression is crucial; many AAC solutions are abandoned due to avoidable interface flaws (see, e.g., Dawe 2006; Fager et al., 2006; Waller 2019). We therefore tried to make clever use of two substitute groups:

- (1) Experts in the field of accessible learning and barrier-free communication, who are able to judge the simplicity/adequacy of the individual steps to be performed; and
- (2) L2 learners, ranging from those with limited computer skills, who reviewed early versions of the interface in order to make the UI as simple as possible, to those with high computer skills, who reviewed the most recent interface to identify the situations in which users expect more proactive linguistic support.

The testing of the newest version of the system with participants with intellectual/learning disabilities had to be suspended for more than a year due to the COVID-19 pandemic. At the time of writing, we have just begun a series of tests (see **Section 4**).

In the following sections, we summarize the main findings<sup>9</sup>.

<sup>9</sup>Additional details on the test setups can be found in the **Supplementary Material** of this paper.

#### 3.6.1 System Evaluation via Interviews With AAC Experts

The group of LS readers is very heterogeneous (Bredel and Maaß, 2016, p. 139). We talked to experts in AAC and LS familiar with its diverse needs. The expert group consisted of:

- (1) A male LS reader with learning disabilities, who regularly reviews LS texts and does some writing himself; accordingly, he can be qualified as an (advanced) real user from our target group;
- (2) A male LS writer, who leads an LS writing workshop and regularly writes and proofreads LS texts;
- (3) A female AAC expert, who implements AAC solutions within a facility for people with multiple and/or severe cognitive or sensory disabilities; and
- (4) A male domain expert, who has worked with people with severe cognitive and physical disabilities for over 35 years and is familiar with many communication methods and their evolution.

We used the *case study* method<sup>10</sup>, employing exploratory think-aloud probes followed by semi-structured interviews. As the first probe, the participants were asked to write a series of prepared sentences of varying complexity, then, to freely formulate a text with *ExtendedEasyTalk*. In the interview, we focused on the participants' assessment regarding potential reasons for the (non-)acceptance of our system. Important insights are presented below.

All experts acknowledged that *ExtendedEasyTalk* meets the requirements of users, who know alphabetic characters and have basic spelling skills but have difficulties writing whole words or complex sentences and coherent texts.

All experts gave positive feedback on the use of AAC symbols in combination with word forms in *ExtendedEasyTalk*. They

<sup>10</sup>The qualification of this study as a case study results from its matching the four key criteria given by Lazar et al. (2017): (1) In-depth examination of a small number of cases, (2) examination in context, (3) multiple data sources, and (4) emphasis on qualitative data and analysis.

appreciated the fact that the range of both symbols and vocabulary is easily customizable and expandable in *ExtendedEasyTalk* as it is crucial for users to be able to use familiar vocabulary and proper names to describe their everyday lives and individual interests. One of the AAC experts phrased it like this: “People need to be able to describe their world in their own words.” This includes users being able to use familiar symbols (e.g., PCS or METACOM<sup>11</sup>), given that AAC symbols and words are commonly learned in combination. In the broader area of evaluation for AAC needs, they focused on the accessibility of *ExtendedEasyTalk*. The experts appreciated the read-aloud function for the produced text in *ExtendedEasyTalk* and the option for users to use their own input devices (e.g., accessible keyboards). They suggested adding the ability to operate the system by *scanning*<sup>12</sup>, i.e., the system iterates sequentially through all options until the user instructs the system to stop and make a selection.

The concept of writing a sentence by answering sequences of wh-questions was appreciated by all experts. They related the answering of wh-questions to parent–child dialogues (see, e.g., Brandt et al., 2016). The experts liked the predictive force of the completion list, which reduces the need for typing and supports correct inflection. They anticipate that this will simultaneously give users a feeling of security and speed up typing. Regarding the mechanism for using complement-taking verbs, we received positive feedback from our LS reader and the AAC expert; they described it as a “reasonable way” to access these constructions. However, all experts recommended thorough testing with varied groups of LS users. The RST-related aspects of the system were recognized by the experts as a good way to practice connecting sentences.

We were pleased with the largely positive feedback and the high level of interest in the system shown by the experts; however, we are aware that the participants knew they were talking to the developer. Moreover, we keep in mind the general warning that one should not over-generalize case study results.

### 3.6.2 Testing With Two L2-learner Groups

#### *Testing with L2 learners with low computer and low literacy skills*

Here, we again chose the case study method as an appropriate way to gain insights. We tested the system with three male L2 learners with predominantly oral German language skills at Common European Frame of Reference (CEFR) L2-level A1–A2<sup>13</sup>. All three are literate in their native languages (Amharic, Tigrinya, and French/Cotocoli). Their computer skills were rudimentary. They were able to write only very short messages in German, e.g., to make appointments *via* messenger apps.

We conducted semi-structured interviews (in ELS wording) assessing the supportive features for sentence

formulation—supplemented with situational follow-up questions to evaluate possible workarounds when deficiencies were discovered. Here, we sum up findings of interest that have led to revisions in the interface.

During all tests, the same barrier to selecting word forms from the completion list hindered fast typing: Spelling errors or mistakes in selecting the gender of an article (*der/die/das*<sub>inflected</sub> ‘the’) resulted in unexpected completion lists. The support while entering sentences with complement-taking verbs was highly appreciated. (*ExtendedEasyTalk* automatically moves the infinitive (with *zu*) to the clause-final position in German—a different position from the participants’ mother tongues.) Moreover, all participants liked the support for correctly conjugating verbs and choosing correctly inflected word forms.

Without over-generalizing, we observed that beginner-level users neither recognize nor use the full extent of the system’s linguistic scope and support. This observation led to the design of an active teaching strategy in form of the “writing workshop” (cf. Section 3.5). At the same time, it demonstrated that users are able to write according to their personal preferences and skill levels. We plan to document in a longitudinal study whether users improve their personal writing skills over time with the support of *ExtendedEasyTalk*. Finally, we received positive feedback regarding the combination of words and visual symbols. Participants emphasized that this helped them recognize and remember words more easily. Without being asked in the probes, the users actively resorted to the read-aloud function.

Based on these observations, we developed the current interface of *ExtendedEasyTalk*, presented in this study.

#### *Testing with L2 learners with high computer and high literacy skills*

In recent test sessions via remote desktop control, conducted with ten L2 learners with different native languages (Arabic, Romanian, Swedish, Mandarin, and Spanish), we tried to find indications of the linguistic support our target group would expect from an advanced writing tool. We recruited this group from among IT experts with German skills between CEFR L2-level A1–B2; additionally, all participants were fluent in English. This was therefore used as the common language between participants and the interviewer during test sessions.

Here, we apply *discount testing* (Nielsen, 1989), a well-established method in *Human–Computer Interaction* (HCI), which was essentially born of necessity. Usability projects should not fall at the hurdle of small budgets. To avoid this, Nielsen proposed a methodology for cheap usability testing: With a handful of participants, a focus on qualitative studies, and the use of the think-aloud method, the majority of usability flaws can be identified. We have chosen the term “discount testing” rather than “case study” here to emphasize that the tasks to be performed were fixed, in contrast to those in the two earlier evaluations (reported above), which allowed for digression during probing (cf. condition (2) in Footnote 11; we are aware that the line between case studies and discount testing is blurred.)

<sup>11</sup>[www.goboardmaker.com/pages/picture-communication-symbols](http://www.goboardmaker.com/pages/picture-communication-symbols) (PCS symbols) and [www.metacom-symbole.de](http://www.metacom-symbole.de) (METACOM symbols).

<sup>12</sup><https://praacticalaac.org/tag/scanning/> (broad overview of different scanning techniques).

<sup>13</sup>[www.coe.int/en/web/common-european-framework-reference-languages](http://www.coe.int/en/web/common-european-framework-reference-languages) (CEFR).

**TABLE 10** | Table of all abbreviations.

Systems/Theories/ Corpora	Expansion
AAC	Augmentative and Alternative Communication
ELS	Extended Leichte Sprache
LS	Leichte Sprache
LST	Leichte Sprache treebank
NLG	Natural-language generation
RST	Rhetorical Structure Theory
TüBa-D/S	treebank of spoken German (systematically called VERBMOBIL)
TüBa-D/Z	treebank of written German
VERBMOBIL	synonym for TüBa-D/S
UI	user interface
VUI	voice user interface
Linguistic terms	Expanded term
DO	direct object
IO	indirect object
NP	noun phrase
POS	part-of-speech (see next panel for POS tags used here)
PP	prepositional phrase
PRED	predicate
SB	subject
V2	verb second word order in German main declarative clauses
VF	verb final word order in German subordinate clauses
SVO	the canonical word order in German main declarative clauses
Used POS tags	Part of Speech
APPO	post-positioned preposition
APPRART	pre-positioned prepositions with agglutinated definite determiner
APPR	pre-positioned preposition
APZR	circum-positioned preposition
KOUI	subordinating conjunction followed by <i>um zu</i> + INFinite
KOUS	subordinating conjunction followed by a sentence
NN	lexical noun
PDS	substituting demonstrative pronoun
PIS	substituting indefinite pronoun
PPER	irreflexive personal pronoun
PPOSS	substituting possessive pronoun
PRELS	relative pronoun
PRF	reflexive personal pronoun
PWS	substituting interrogative pronoun
SVP	separable verb prefix
VA(FIN/IMP/PP/INF/IZU)	auxiliary as FINite form/IMPerative/PastParticiple/ INFinitive/InfwithZU
VM(FIN/IMP/PP/INF/IZU)	modal as FINite form/IMPerative/PastParticiple/INFinitive/ InfwithZU
VV(FIN/IMP/PP/INF/IZU)	lexical verb as FINite form/IMPerative/PastParticiple/ INFinitive/InfwithZU
POS wild cards	Expanded term
APPR,*	APPR and APPRART (“*” refers to zero to <i>n</i> arbitrary characters)
N.*	NE and NN
PRO	PDS and PIS and PPOSS and PRELS and PRF and PPER and PWS
V.FIN	VAFIN/VMFIN/VVFIN (“*” refers to one arbitrary character)

To obtain comparable results, we presented each participant with a picture supplemented by five sentences that provided the necessary German vocabulary, allowing the participants to focus on typing rather than on finding German words. We asked them to think aloud while typing sentences in *ExtendedEasyTalk* and using the

connectors provided in Panel 2 (B) to relate the sentences in a short story—supposedly to be emailed to a solely German-speaking child. At the end of each session, we conducted a short semi-structured interview and asked the users to fill in a user experience questionnaire<sup>14</sup>, in English for their convenience. We focused mainly on the question: What support do IT specialists with few (written) German skills expect from an advanced writing system? We also took notice of any flaws in the menus.

The results for within-sentence support were mainly positive; all participants felt supported by the system. Moreover, they reported that they found the system easy to understand and intuitive to operate. This feedback is reflected in the fact that all learners managed to use the system autonomously, following a demonstration in which the interviewer produced an example sentence using *ExtendedEasyTalk*. As in the other L2-learner group, the IT experts appreciated the combination of symbols and words, the automatic inflection of words, and the word-ordering support.

However, this group asked for active support in spelling and/or finding a German word—an issue we also noticed in the earlier test sessions. This problem needs to be addressed in the next *ExtendedEasyTalk* prototype, particularly as many LS readers presumably have spelling problems as well.

We want to take up the idea of three participants who proposed adding suggestions based on words entered via a microphone to the already available vocabulary suggestions based on spelling. We are planning a sub-series of tests with users with functional speech, employing a speech recognition device as a voice user interface (VUI). In line with suggestions by our participants, we want to implement a mode offering lemmas for a preselected topic or domain—similar to common AAC grid layouts where the vocabularies of customizable categories or contexts, such as ‘food’, ‘hygiene’, ‘in school’, or ‘at home’, are proposed (cf. the AAC systems presented in **Section 3.1.1**).

The menu producing sentence connectors outlined in **Section 3.4** was judged to be intuitive and meaningful by all participants. However, all participants expected the system to provide feedback on the quality of their choices. This expectation probably originates from the fact that most participants have experience with fill-the-gap German language-learning software. While they certainly appreciated the freedom that *ExtendedEasyTalk* offers compared to such exercises, they would welcome help in this area similar to the system’s word order expertise. This problem is difficult to solve in *ExtendedEasyTalk* as we have no overall content representation to determine what is a reasonable/necessary relation to use.

The system was perceived by all participants as uncluttered, organized, and clearly structured. However, the IT experts criticized the “outdated” look and feel of its graphical UI design. We assume that this feedback is connected to the fact that people working in IT are used to modern UI aesthetics<sup>15</sup>, whereas the UI of *ExtendedEasyTalk* is oriented towards the design patterns of AAC applications and focuses on accessible

<sup>14</sup><https://www.ueq-online.org/> (user experience questionnaire UEQ).

<sup>15</sup><https://developer.apple.com/design/> (“Flat Design”); <https://material.io/design/> (“Material Design”).

design and good readability for the target group (cf. the UI design of the AAC systems cited in **Section 3.1** and the style guide for LS texts in Netzwerk Leichte Sprache).

Nevertheless, according to the user-experience questionnaire, eight of the 10 participants would recommend or use the system for language learning. Six of them would prefer to run *ExtendedEasyTalk* as a smartphone app instead of a desktop app. In line with the recent developments in AAC technology on common commercial handheld devices (which we have cited above), a new version of *ExtendedEasyTalk* for smartphones and tablets is on our to-do list.

## 4 CONCLUSIONS AND FUTURE WORK

We have defined Extended Leichte Sprache (ELS), a native extension of LS based on observations from a corpus study of LST, VERBMOBIL, and TüBa-D/Z. *ExtendedEasyTalk* facilitates fast and correct typing for the target language, ELS, by employing linguistic processing to a large extent. The system strives for correct and understandable text during both sentence production and sentence combination. Interactive user guidance is tailored to the personal level of grammatical support needed to produce correct and coherent complex content. Most importantly, linguistic decisions are formulated in an easy and intuitive manner. In order to promote LS text production by the LS community themselves, *ExtendedEasyTalk* teaches text-writing skills to more advanced users. In the writing-workshop mode, the writer is trained to foresee/resolve underspecifications that result from matching the presented facts with the readers' presupposed knowledge.

As mentioned above, usability tests with the target group in the presence of the researchers have been impossible for more than a year now due to COVID-19. Unfortunately, people with disabilities are particularly affected by the social isolation caused by COVID-19 (Rödler, 2020; Portal et al., 2021). This highlights the importance of the further development of systems like *ExtendedEasyTalk* that support the target group of LS readers to communicate their thoughts (remotely) in the form of text (messages). We are currently conducting a broad-scope system-evaluation study with members of the target population, aiming to gain new insights to guide the redesigning of the UI for the next version of *ExtendedEasyTalk*. As mentioned in the evaluation in **Section 3.6.2**, this prototype should include a better strategy in Panel 3 to avoid empty suggestion lists when choosing words. A series of prototypes for a cell phone version is also under evaluation.

In addition, we would like to extend the system's writing-workshop mode. For example, in a revision phase across the text provided in Panel 1, referents could become pronominalized or elided in consecutive SVO sentences. The use of other types of substituting pronouns (e.g., demonstratives [PDS]) could be trained. We realize that these constructions are not authorized by all LS definitions; however, they are frequent in LS texts, and readers can therefore be expected to be familiar with them. As pronouns shorten a text, writers should be trained to use them.

Given the fact that our natural-language paraphrase generator administers the syntactic structures of all typed sentences, we

plan to implement an additional export function that transforms the typed LS text into syntactic constructions in standard German. So, documents for non-LS readers can automatically be produced. For this purpose, a component similar to the Aggregator (component (2) in the NLG blue print in **Figure 1**) has to be added. As mentioned, the difficulty with this task is the problem to map appropriate portions of sentences and their connectors onto the overall discourse structure of the text. The latter topic is related to the question we brought up in **Section 2.3**. What is a good way of replacing standard German constructions by (E)LS conform ones? For finding easy-to-perform rules of thumb, we plan more thorough studies into LST.

## SORTED LIST OF ABBREVIATIONS USED IN THE PAPER

Due to the interdisciplinary nature of our study, in **Table 10**, we provide a list of all acronyms used in the text, grouped in four panels.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

Ethical review and approval were not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

This is joint work; the substantive focus of each author was as follows: KH: ELS definition based on corpus study; and IS: *ExtendedEasyTalk* prototype and evaluations.

## ACKNOWLEDGMENTS

The authors are extremely grateful to the two reviewers and the editor Scott Hale for their constructive, very detailed, and insightful suggestions and comments. All remaining errors are the authors' own responsibility.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcomm.2021.689009/full#supplementary-material>

## REFERENCES

- Bell, A. (1984). Language Style as Audience Design. *Lang. Soc.* 13, 145–204. doi:10.1017/S004740450001037X
- BITV2.0 (2011). Verordnung zur Schaffung Barrierefreier Informationstechnik Nach dem Behindertengleichstellungsgesetz (Barrierefreie-Informationstechnik-Verordnung - BITV 2.0). Barrierefreie-Informationstechnik-Verordnung vom 12. September 2011 (BGBl. I S. 1843), die zuletzt durch Artikel 1 der Verordnung vom 21. Available at: [http://www.gesetze-im-internet.de/bitv\\_2\\_0/BjNR184300011.html](http://www.gesetze-im-internet.de/bitv_2_0/BjNR184300011.html).
- Bock, B. M. (2017). Das Passiv- und Negationsverbot 'Leichter Sprache' auf dem Prüfstand – Empirische Ergebnisse aus Verstehenstest und Korpusuntersuchung. *Sprachreport*. 33 (1), 20–28. doi:10.1007/978-3-476-04852-3\_39
- Bock, B. M. (2019). *Leichte Sprache* – Kein Regelwerk: Sprachwissenschaftliche Ergebnisse und Praxisempfehlungen aus dem LeISA-Projekt. Berlin, Germany: Frank & Timme.
- Brandt, S., Lieven, E., and Tomasello, M. (2016). German Children's Use of Word Order and Case Marking to Interpret Simple and Complex Sentences: Testing Differences Between Constructions and Lexical Items. *Lang. Learn. Development*. 12, 156–182. doi:10.1080/15475441.2015.1052448
- Bredel, U., and Maaß, C. (2016). *Leichte Sprache: Theoretische Grundlagen. Orientierung für die Praxis*. Berlin, Germany: Bibliographisches Institut.
- Dawe, M. (2006). "Desperately Seeking Simplicity," in Proceedings of the 2006 Conference on Human Factors in Computing Systems (SIGCHI 2006), Montréal, QC, Canada, April 22–27, 2006, 1143–1152. doi:10.1145/1124772.1124943
- Demasco, P. W., and McCoy, K. F. (1992). Generating Text From Compressed Input. *Commun. ACM*. 35, 68–78. doi:10.1145/129875.129881
- Fager, S., Hux, K., Beukelman, D. R., and Karantounis, R. (2006). Augmentative and Alternative Communication Use and Acceptance by Adults With Traumatic Brain Injury. *Augmentative Altern. Commun.* 22, 37–47. doi:10.1080/07434610500243990
- G2.com (2021). Best Natural Language Generation (NLG) Software. Available at: <https://www.g2.com/categories/natural-language-generation-nlg> (Accessed January 16, 2021).
- Gatt, A., and Krahmer, E. (2018). Survey of the State of the Art in Natural Language Generation: Core Tasks, Applications and Evaluation. *J. Artif. Intelligence Res.* 61, 65–170. doi:10.1613/jair.5477
- Gulikers, L., Rattnik, G., and Piepenbrock, R. (1995). *German Linguistic Guide of the CELEX Lexical Database Tech. Rep.* Philadelphia, MA: Linguistic Data Consortium.
- Hansen-Schirra, S., and Maaß, C. (2020). *Easy Language Research: Text and User Perspectives*. Berlin, Germany: Frank & Timme. doi:10.26530/20.500.12657/42088
- Harbusch, K., Härtel, J., and Cameron, C. J. (2014). "Automatic Feedback Generation for Grammatical Errors in German as Second Language Learners Focused on the Learner's Personal Acquisition Level," in Proceedings of EUROCALL 2014 - Call Design: Principles and Practice (Groningen, Netherlands), 140–145.
- Harbusch, K., and Kempen, G. (2002). "A Quantitative Model of Word Order and Movement in English, Dutch and German Complement Constructions," in Proceedings of the 19th International Conference on Computational Linguistics – Volume 1 (COLING '02), Taipei, Taiwan, 1–7. doi:10.3115/1072228.1072262
- Harbusch, K., and Kempen, G. (2011). "Automatic Online Writing Support for L2 Learners of German Through Output Monitoring by a Natural-Language Paraphrase Generator," in WorldCALL - International Perspectives on Computer-Assisted Language Learning, New York, NY: Routledge/Taylor&Francis Group, 128–143.
- Harbusch, K., and Kempen, G. (2007). "Clausal Coordinate Ellipsis in German: The TIGER Treebank as a Source of Evidence," in Proceedings of the Sixteenth Nordic Conference of Computational Linguistics (NODALIDA 2007), Tartu, Estonia, 81–88.
- Harbusch, K., and Kempen, G. (2009). "Coordinate Ellipsis and its Varieties in Spoken German: A Study With the TüBa-D/S Treebank of the VERBMOBIL Corpus," in Proceedings of the 8th International Workshop on Treebanks and Linguistic Theories (TLT 8), Milano, Italy, 83–94.
- Harbusch, K., van Breugel, C., Koch, U., and Kempen, G. (2007). "Interactive Sentence Combining and Paraphrasing in Support of Integrated Writing and Grammar Instruction: A New Application Area for Natural Sentence Generators," in Proceedings of the 11th European Workshop on Natural Language Generation (ENLG 2007), Dagstuhl, Germany, 65–68.
- Henry, S. L. (2007). *Just Ask: Integrating Accessibility Throughout Design*. Madison, WI: Shawn Lawton Henry. Available at: [www.uiAccess.com/JustAsk/](http://www.uiAccess.com/JustAsk/).
- Hovy, E. H. (1988). "Planning Coherent Multisentential Text," in Proceedings of the The 26th Annual Meeting of the Association for Computational Linguistics (ACL 88), New York, NY, 163–169. doi:10.3115/982023.982043
- Inclusion Europe (2009). Informationen für alle – Europäische Regeln, wie man Informationen leicht lesbar und leicht verständlich macht. Available at: [https://www.inclusion-europe.eu/wp-content/uploads/2017/06/DE\\_Information\\_for\\_all.pdf](https://www.inclusion-europe.eu/wp-content/uploads/2017/06/DE_Information_for_all.pdf) (Accessed March 29, 2021).
- Kempen, G., and Harbusch, K. (2019). Mutual Attraction Between High-Frequency Verbs and Clause Types With Finite Verbs in Early Positions: Corpus Evidence From Spoken English, Dutch, and German. *Lang. Cogn. Neurosci.* 34, 1140–1151. doi:10.1080/23273798.2019.1642498
- Kempen, G., and Harbusch, K. (2002). Performance Grammar: A Declarative Definition. *Lang. Comput.* 45, 148–162. doi:10.1163/9789004334038\_013
- Kempen, G., and Harbusch, K. (2016). Verb-second Word Order After German Weil 'Because': Psycholinguistic Theory From Corpus-Linguistic Data. *Glossa: A J. Gen. Linguistics*. 1 (1), 1–32. doi:10.5334/gjgl.46
- König, E., and Lezius, W. (2003). *The TIGER Language: A Description Language for Syntax Graphs, Formal Definition. Tech. Rep.* Universität Stuttgart: Institut für maschinelle Sprachverarbeitung (IMS).
- Kuhn, T. (2014). A Survey and Classification of Controlled Natural Languages. *Comput. Linguistics*. 40, 121–170. doi:10.1162/COLI\_a\_00168
- Lancioni, G. E., Singh, N. N., O'Reilly, M. F., and Alberti, G. (2019). Assistive Technology to Support Communication in Individuals With Neurodevelopmental Disorders. *Curr. Dev. Disord. Rep.* 6, 126–130. doi:10.1007/s40474-019-00165-x
- Lazar, J., Feng, J. H., and Hochheiser, H. (2017). *Research Methods in Human Computer Interaction*. 2nd Ed Cambridge, MA: Morgan Kaufmann, an imprint of Elsevier. doi:10.1016/B978-0-12-805390-4.00016-9
- Lee, D.-S., and Yoon, W. C. (2004). Quantitative Results Assessing Design Issues of Selection-Supportive Menus. *Int. J. Ind. Ergon.* 33, 41–52. doi:10.1016/j.ergon.2003.07.004
- Lieske, C., and Siegel, M. (2014). Verstehen Leicht Gemacht. *Technische Kommunikation*. 1, 44–49.
- Light, J., McNaughton, D., and Caron, J. (2019). New and Emerging AAC Technology Supports for Children With Complex Communication Needs and Their Communication Partners: State of the Science and Future Research Directions. *Augmentative Altern. Commun.* 35, 26–41. doi:10.1080/07434618.2018.1557251
- Light, J., and McNaughton, D. (2012). The Changing Face of Augmentative and Alternative Communication: Past, Present, and Future Challenges. *Augmentative Altern. Commun.* 28, 197–204. doi:10.3109/07434618.2012.737024
- Linzen, T., and Baroni, M. (2021). Syntactic Structure from Deep Learning. *Annu. Rev. Linguist.* 7, 195–212. doi:10.1146/annurev-linguistics-032020-051035
- Löffler, C. (2015). Leichte Sprache als Chance zur Gesellschaftlichen Teilhabe Funktionaler Analphabeten. *Didaktik Deutsch*. 38, 17–23.
- Maaß, C., Rink, I., and Zehrer, C. (2014). *Leichte Sprache in der Sprach- und übersetzungswissenschaft. Susanne Jekat, Heike Elisabeth Jüngst, Klaus Schubert & Claudia Villiger (Hg.), Sprache barrierefrei gestalten. Perspektiven aus der Angewandten Linguistik*. 53, 86.
- Mann, W. C., and Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. *Interdiscip. J. Study Discourse*. 8 (3), 243–281. doi:10.1515/text.1.1988.8.3.243
- Netzwerk Leichte Sprache (2013). Die Regeln für Leichte Sprache. Available at: [https://www.leichte-sprache.org/wp-content/uploads/2017/11/Regeln\\_Leichte\\_Sprache.pdf](https://www.leichte-sprache.org/wp-content/uploads/2017/11/Regeln_Leichte_Sprache.pdf) (Accessed March 29, 2021).
- Ney, J. W. (1980). A Short History of Sentence Combining: Its Limitations and Use. *English Education*. 11, 169–177. doi:10.1177/003368828001100103

- Nielsen, J. (1989). "Usability Engineering at a Discount," in Proceedings of the Third International Conference on Human-Computer Interaction on Designing and Using Human-Computer Interfaces and Knowledge Based Systems (USA: Elsevier Science Inc), 394–401. doi:10.5555/92449.92499
- Nordquist, R. (2018). An Introduction to Sentence Combining. ThoughtCo. Available at: <https://www.thoughtco.com/an-introduction-to-sentence-combining-1692421> (Accessed March 18, 2021).
- Nüssli, N. D. (2019). *Übersetzen in die Leichte Sprache: Übersetzungsprobleme, Übersetzungslösungen und Auswirkungen auf das Textverständnis von Menschen mit Downsyndrom: am Beispiel von Texten zum Thema Gesundheit*. Zürich, Switzerland: ZHAW Zürcher Hochschule für Angewandte Wissenschaften. doi:10.21256/zhaw-3344
- Ogden, C. K. (1930). *Basic English: A General Introduction with Rules and Grammar*, Vol. 1. London, UK: Paul Treber&Co.
- Petrova, S., Solf, M., Ritz, J., Chiarcos, C., and Zeldes, A. (2009). Building and Using a Richly Annotated Interlinear Diachronic Corpus: The Case of Old High German Tatian. *Traitement Automatique des Langues*. 50, 47–71.
- Portal, H., Schmidt, G., Fernández, R. C., Marcondes, B., Šveřpa, M., Dragičević, V., et al. (2021). Neglect and Discrimination. Multiplied. How Covid-19 Affected the Rights of People With Intellectual Disabilities and Their Families. Inclusion Europe. Available at: <https://www.inclusion-europe.eu/wp-content/uploads/2020/11/COVID-report-Final.pdf> (Accessed March 09, 2021).
- Pottmann, D. M. (2019). "Leichte Sprache and Einfache Sprache" - German plain Language and Teaching DaF German as a Foreign Language. *Studia Linguistica*. 38, 81–94. doi:10.19195/0137-1169.38.6
- Reichardt, A., Kruse, N., and Lipowsky, F. (2014). Textüberarbeitung mit Schreibkonferenz oder Textlupe. Zum Einfluss der Schreibumgebung auf die Qualität von schülertexten. *Didaktik Deutsch*. 19, 64–85.
- Reis, M. (2013). "Weil-V2"-Sätze und (k)ein Ende? Anmerkungen zur Analyse von Antomo & Steinbach (2010). *Z. für Sprachwissenschaft*. 32, 221–262. doi:10.1515/zfs-2013-0008
- Reiter, E., and Dale, R. (2000). Building Natural Language Generation Systems. *Stud. Nat. Lang. Process.* Cambridge UK, Cambridge Univ. Press. doi:10.1017/CBO9780511519857
- Rödler, P. (2020). Totale Institution. *Bp*. 59, 345–358. doi:10.30820/0341-7301-2020-4-345
- Ross, J. R. (1967). Constraints on Variables in Syntax. Ph.D. Thesis. Cambridge, MA/UASA: MIT.
- Saddler, B., and Preschern, J. (2007). Improving Sentence Writing Ability Through Sentence Combining Practice. *Teach. Exceptional Child*. 29, 6–11. doi:10.1177/004005990703900301
- Sennrich, R., Volk, M., and Schneider, G. (2013). "Exploiting Synergies Between Open Resources for German Dependency Parsing, POS-Tagging, and Morphological Analysis," in Proceedings of the International Conference Recent Advances in Natural Language Processing (RANLP 2013), Hissar, Bulgaria, 601–609.
- Stegmann, R., Telljohann, H., and Hinrichs, E. W. (2000). *Stylebook for the German Treebank in Verbmobil. Tech. Rep.*, 239. Saarbrücken, Germany: DFKI.
- Steinmetz, I., and Harbusch, K. (2020). "Enabling Fast and Correct Typing in 'Leichte Sprache' (Easy Language)," in Proceedings of the The Fourth Widening Natural Language Processing Workshop (WINLP 4), Seattle, CA, USA. 64–67.
- Suter, J., Ebling, S., and Volk, M. (2016). "Rule-based Automatic Text Simplification for German," in Proceedings of the 13th Conference on Natural Language Processing (KONVENS 2016), Zürich, Switzerland, 279–287.
- Telljohann, H., Hinrichs, E. W., Kübler, S., Zinsmeister, H., and Beck, K. (2009). *Stylebook for the Tübingen Treebank of Written German (TüBa-D/Z). Tech. Rep.* Tübingen, Germany: Seminar für Sprachwissenschaft, Universität Tübingen.
- Temperley, D. (2019). Rare Constructions Are More Often Sentence-Initial. *Cogn. Sci.* 43, e12714. doi:10.1111/cogs.12714
- Tritscher, J. (2016). *Evaluation von State of the Art Syntaxparsern auf deutschen Romanen des 16. bis 20. Jahrhunderts*. B.S. Thesis. Würzburg, Germany: Lehrstuhl für Künstliche Intelligenz und Angewandte Informatik an der Universität Würzburg.
- Wahlster, W. (2000). *Verbmobil: Foundations of Speech-To-Speech Translation*. Berlin, Germany: Springer. doi:10.1007/978-3-662-04230-4
- Waller, A. (2019). Telling Tales: Unlocking the Potential of AAC Technologies. *Int. J. Lang. Commun. Disord.* 54, 159–169. doi:10.1111/1460-6984.12449
- Zurstrassen, B. (2015). "Inklusion durch Leichte Sprache? Eine kritische Einschätzung," in *Didaktik der inklusiven politischen Bildung* (Bonn, Germany: Bundeszentrale für politische Bildung), 129–138.

**Conflict of Interest:** We declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Harbusch and Steinmetz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.