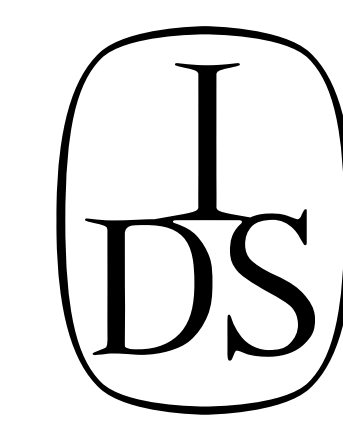


C-WEP – Rich Annotated Collection of Writing Errors by Professionals



INSTITUT FÜR
DEUTSCHE SPRACHE

Cerstin Mahlow <mahlow@ids-mannheim.de>

C-WEP

C-WEP is the Collection of Writing Errors by Professional Writers of German. It currently consists of 245 sentences with grammatical errors. All sentences are taken from published texts. All authors are professional writers with high skill levels with respect to German, the genres, and the topics.

The purpose of this collection is to provide seeds for more sophisticated writing support tools as only a very small proportion of those errors can be detected by state-of-the-art checkers.

C-WEP is annotated on various levels and freely available. C-WEP comes with one or more target hypothesis for each sentence and can thus also be used for the development and testing of grammar checkers.

Error Detection by Grammar Checkers [245 erroneous sentences]

Error Type	System	marked	true pos.	appropriate sugg.	no sugg.	wrong sugg.
name	MS Checker	73	1	0	11	64
	Hunspell	82	1	1	22	59
word	MS Checker	15	5	4	3	8
	Hunspell	47	5	3	29	15
compound issue	MS Checker	6	0	0	0	6
	Hunspell	1	0	0	0	1
grammar issue	MS Checker	59	13	0	0	59
	Hunspell	0	0	0	0	0
grammar as spelling	MS Checker	7	7	6	0	1
	Hunspell	0	0	0	0	0

Error detection by MS Checker and Hunspell. (Names include names for persons, places, and products.)

C-WEP Error Typology

Type	Number	POS involved	Example
missing words	62	all, including phrases and clauses	[1] * Die Römer hatten es im Laufe ihrer Geschichte mehrmals gegnerischen Kriegselefanten zu tun. [2] * Vor anderthalb Jahre starb Sohn des Ehepaars im Alter von 16 Jahren während eines Familienurlaubs. [3] * Außerdem muss der Klub im nächsten Spieler-Draft auf zwei Ziehungen – auf eine in der ersten und eine in der dritten Runde.
duplicate words [discontinuously]	46	all, including phrases	[4] * Wir fliegen wie geplant am 30. Oktober nach Kairo fliegen .
agreement	42	particle + verb, verb + noun, noun phrase	[5] * Der Verband unabhängiger Schweizer Hochzeitsplaner [VSUH] bietet einen Diplomlehrgang zum Hochzeitsplaner durch .
word order	28	complex phrases, noun constructions	[6] * Von der grauhaarigen Frau sieht Christin nur zuckenden den Rücken . [7] * Versuchen Sie nicht, Ihren zu Partner verändern .
duplicate POS [non-identical words]	25	preposition, verb	[8] * Sie sind in auf einem fremden Netz, es gelten die folgenden max. Gebühren [in CHF] [9] * Beide Parteien, heißt es aus Lausanne weiter, haben hätten bereits je einen Vertreter aus dem geschlossenen Schiedsrichterpool des Cas benannt.
duplicate words [continuously]	11	all, including phrases	[10] * Ob es den den Jungen wirklich gibt oder ob er nur eine Phantasiegestalt ist, spielt keine Rolle. [11] * Kurz nach dem nach dem Abitur traf ich meinen ehemaligen Direktor im Pressecafé in der Friedrichstraße.
semi-duplicate words	3	noun compounds, particle verbs	[12] * Der Brieffund bedeutet viel für die Reformations forschung Forschung .

C-WEP Structure

- XML file with DTD
- evidence (+ annotation)
- source
- one or more target hypotheses (+ annotation)
- error category

C-WEP Annotation

- Part of speech tagging [Mbt trained on TüBaD/Z]
- Detailed structured morphological analyses [Stripey Zebra, rule-based]
- Dependency trees [MATE]
- LFG analyses [IMS German ParGram Grammar]

Institut für Deutsche Sprache
R 5, 6-13
68161 Mannheim
Germany
<http://www.ids-mannheim.de>

<http://lingued.info/linguistic-resources/cwep/>

Member of the

Leibniz Association