Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Authenticity of scientific literature – Analyzing the usage of hedging in adapted scientific literature

Hilko Aljets
University of Göttingen
Thomas Waitz
twaitz@gwdg.de
University of Göttingen

Research Article

Keywords: natural language processing, hedging, adapted scientific literature, scientific literature, naive bayes, support vector

Posted Date: May 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9266058/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientometrics on September 2nd, 2026.
See the published version at https://doi.org/10.1007/s11192-026-05785-w.

Authenticity of scientific literature – Analyzing the usage of hedging in adapted scientific literature

Authors and Affiliations

- *Dr. Hilko Aljets, University of Göttingen, Institute of Inorganic Chemistry, Department of Chemistry Education, Tammannstr. 4, D-37077 Göttingen, Germany, Email: hilko.aljets@uni-goettingen.de, https://orcid.org/0000-0001-7464-0535
- Prof. Dr. Thomas Waitz, University of Göttingen, Institute of Inorganic Chemistry, Department of Chemistry Education, Tammannstr. 4, D-37077 Göttingen, Germany, Email: twaitz@gwdg.de, https://orcid.org/0000-0002-4098-1224

*Corresponding author (Hilko Aljets)

Abstract

Adapted scientific literature (ASL) is increasingly used in educational contexts to make authentic science texts accessible to students. One hallmark of scientific writing is the use of hedging—linguistic means to express uncertainty or tentativeness. This study investigates whether ASL maintains this characteristic feature, and to what extent its usage of hedging aligns with original scientific articles. Using a machine learning approach, we trained classifiers on the BioScope corpus to automatically detect hedging in English and German texts. We then analyzed corpora consisting of original research articles, their adapted counterparts, and textbook excerpts. Our results show that ASL retains the typical distributional pattern of hedging—particularly its increased frequency in the introduction and discussion sections—but the overall proportion of hedging expressions is reduced compared to the original texts. In contrast, textbook texts contain very little hedging and present scientific knowledge as more certain. These findings suggest that ASL offers a linguistically authentic alternative to original scientific literature and may help learners engage with the epistemic practices of science more realistically. While some reduction in

hedging occurs during the adaptation process, ASL preserves essential rhetorical features of scientific discourse and can thus serve as a valuable resource in science education.

Keywords: natural language processing; hedging; adapted scientific literature; scientific literature; naive bayes; support vector

Introduction

In science education, reading and understanding scientific literature offers the potential to gain insights into the way scientists think and work (Fang, 2005; Phillips & Norris, 2009; Yarden, 2009). However, reading original scientific literature is challenging for students due to its highly specialized language and content, which is why comprehension of the text suffers. The concept of adapted primary literature (APL) by Yarden et al. (2015) aims to adapt the content of original scientific literature to the prior knowledge of the respective learning group while retaining important structural characteristics, possibly promoting the development of an understanding of science. This would enable learners to get in touch with one of the essential communications tools in science in a authentic way, not only disseminating content knowledge but also meta-aspects of science like nature of science (I. Braun & Nückles, 2014; Hagen et al., 2022).

One underlying premise of adapted scientific literature (ASL) is the authentic representation of scientific literature. Systematic research examining the linguistic properties of ASL and the corresponding original literature is missing until now. To our knowledge, only Ariely et al. (2019) analyzed linguistic features of one adapted primary literature, the original research article and a popular article. A total of five linguistic features were analyzed by the authors: informational density, abstraction, technicality, authoritativeness, and hedging. They found a lowered lexical complexity for the APL article, meaning a lesser use of technical terms, and more use of verbs instead of nominalizations. However, the found hedges in the research article and the APL Article had similar rhetorical functions. Overall, the authors concluded that the adaptation leads to a lower lexical complexity while staying authentic to scientific writing.

With this study, we want to further examine the authenticity of ASL by analyzing the use of hedging on a broader database. In contrast to Ariely et al., we focus on the quantity and distribution of hedges instead of analyzing their lexical function. Specifically, we want to explore if the quantity of hedged statements is similar between the examined genres and if the adaptation conserves the distribution of hedges within the text progression? For this, we use a machine learning approach, since classifying all sentences by hand is a lot of work. Furthermore, after establishing our method, it is easy applicable to new articles and texts.

Scientific literature

Science communication is an expression of the social character of science and constitutes a diverse field with many different communication methods. Focusing on written science communication, Goldman and Bisanz (2002) distinguish three major sociocultural roles for disseminating scientific information: (1) communicating among scientists, (2) popularizing information to the general public and (3) providing formal education. Each of these roles is characterized by distinct forms of communication with characteristic properties, tailored to the needs of the target audience. Fig. 1 provides a model of the main sociocultural roles and their dominant text genres, complemented by the new genre of ASL.

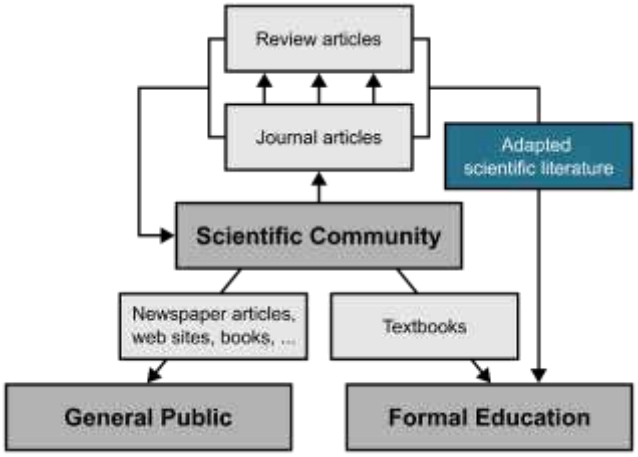


Fig. 1 Overview of the major sociocultural roles in science communication and their mainly used text genres. The illustration is adapted from Authors (2023) [names deleted to maintain the integrity of the review process], theoretically based on the work of Goldman and Bisanz (2002) and augmented by the concept of adapted scientific literature. Originating in the scientific community, scientific knowledge is communicated to scientists, popularized to the general public, and provided for the formal education through specific text genres like journal and review articles, newspaper articles and textbooks, respectively. Combined under the umbrella term adapted scientific literature, adapted primary literature (APL) and adapted review articles (ARA) are new hybrid genres, which are reduced forms of scientific literature and are directed to the formal education (Authors, 2023 [names deleted to maintain the integrity of the review process]; (Yarden et al., 2015).

With the term ASL we want to subsume adaptations of scientific literature, and therefore refer to APL described by Yarden et al. (2001, 2015), which mainly focuses on journal articles, as well as to adaptations of review articles, which we denote as adapted review articles (ARA). Adapted scientific texts represent modified versions of original scientific texts like research or review articles. Goal of the adaptation process is to conserve the structural and rhetorical features of the original texts while making them more accessible to students (Yarden et al., 2001, 2015).

Original scientific literature assumes a high degree of prior knowledge from the reader and relies heavily on background knowledge within a specific scientific field. Even for scientists from neighboring fields of research, reading articles outside their domain can be challenging (Bazerman, 1985). This difficulty is amplified for students who lack experience with the text structures and are confronted with a dense array of details and unfamiliar research methods (Goldman & Bisanz, 2002; Yarden et al., 2001). The adaptation process consists of several steps: removing details from the methods and the results section, expanding the introduction section with necessary background information, and adding context to the discussion section, rendering the article more understandable for novices (Yarden et al., 2001, 2015).

In a previous article we already established the ASL term (Authors, 2023 [names deleted to maintain the integrity of the review process]). A more detailed paper on the theoretical foundation on the concept is currently in preparation.

Hedging

Conveying uncertainty is one elemental part of each scientific publication – either in form of a proper error propagation or in form of a specific choice of wording. This is reflected in science communication: by using specific words or phrases "whose job it is to make things fuzzier or less fuzzy" (Lakoff, 1973, p. 471), the epistemic strength of the statement is weakened. These words and phrases are called hedges and their use is essential for the natural sciences (Hyland, 1994, 1996). There are many ways to manipulate the certainty of a statement: by using verbs (e.g., The data *indicates…*), adverbs (e.g., It is *unclear*…) or discourse strategies (e.g., *It is difficult to conclude*…). Hyland also observed a clustering of hedges in one sentence (e.g., *this would probably suggest*), reinforcing the strength of the modification. Some common hedges for the English language collected by Hyland (1996) and Chen et al. (2018) are listed in Table 1.

Table 1 Typical hedges from research articles collected from Hyland (1996) and Chen et al. (2018)

Hyland (1996)	**Chen et al. (2018)**
would (not), may (not), could, might (not), should, cannot, will (not), must, shall, ought to, indicate, suggests, appear, propose, seem, report, predict, assume, ...	unclear, controversial, inconclusive, inconsistent, confusing, uncertain, uncertainty, unknown, ambiguous, incomplete, contradictory, ...

By weakening the strength of the scientific statement, authors use hedging to defend themselves against claims made by the readers, i.e., the scientific community. Like other rhetorical figures, hedging is not some sort of grammatic. Rather, hedging is constructed within the author-reader relationship, based on the circumstances of the communication (Hyland, 1994, 1996; Schröder, 1998). Studies from Lewin (2005) and Hampp (2022) indicate a divergence in the perception of hedging between readers and authors of scientific publications. Not only do they identify hedges differently, but readers and writers also report different motivations for the use of hedges, both specifically and generally. This shows that the analysis of hedging only on a pure text level is not sufficient. Since readers and authors can understand hedges differently, hedging needs also to be examined from the perspective of those participating in the communication.

The extent to which hedging is perceived consciously or subconsciously is unclear. Studies from Hyland (2000), Low (1996), and Takimoto (2015) suggest that hedging may be "lexically invisible" for students, while boosters (expressions strengthening the epistemic power of a statement, e.g., *definitely*, *clearly*) were almost always perceived as such. However, the results are only somewhat generalizable, as the respondents in the studies by Hyland and Takimoto were asked to evaluate a scientific text in English and English was not their first language. In contrast to this are the results of Crismore and Vande Kopple (1997a, 1997b), who compared the effects of reading texts with and without hedging: They found that readers' attitudes towards the content of the texts changed more positively when the text contained hedges. Similarly, Jensen (2008) found that journalists of news articles on scientific research as well as the scientists involved are rated as more credible and trustworthy if they include hedging. Taken together, this suggests that hedging has a more subconscious effect on readers. Despite its complex effect on readers, hedging remains one of the most paramount features of scientific writing.

Aim of study

The presented study aims to explore the use of hedging in original an adapted scientific writing. In particular, we are interested in whether adapted articles deal with uncertainty in a similar way to their originals. This question is two-folded: (1) Can hedging be found in a similar quantity between the genres in question? (2) Is the heading in adapted articles similarly distributed to the original articles and can a higher density be found in the introduction and discussion sections? Since Airely et al. (2019) already showed a functional comparability between hedging in adapted and original articles, we will focus on the quantity and presence of hedging.

Method

In order to follow our research questions, we used a machine learning approach to classify each sentence of our analysis objects according to the presence of hedging.

Corpora

Reference dataset

As a reference dataset we used the freely available BioScope corpus from Vincze et al. (2008), which consists of three parts: medical reports, biological full papers and biological scientific abstracts. Unfortunately, the medical reports could no longer be downloaded, so we only used the full papers and abstracts (Table 2). However, using medical reports for our application would also be arguable, as they differ in nature from classical scientific articles. The dataset is annotated by two independent linguists who have tagged negative and speculative keywords at the token level and their linguistic scope at the sentence level, and can be downloaded as XML files (see Fig. 7, Appendix/SI for an excerpt). In our analysis, we focused on the speculative keywords.

Table 2 Components of the BioScope corpus (Vincze et al., 2008). The corpus is composed of two subcorpora: full papers and abstracts. The original corpus also includes a further sub-corpus on public medical reports which, however, is no longer available.

	Full Paper	**Abstracts**	**Total**
n	9	1,273	-
Sentences	2,670	11,871	14,541
Sentences with hedging / %	19.4	17.7	18.0

As the adapted articles we used for our study were prepared in German, a reference dataset was generated by translating the dataset using DeepL[1]. The quality of the translation was checked randomly. In a similar approach, Sifa et al. (2019) showed a good match between the well-known Stanford Natural Language Inference (SNLI) dataset and a translated version they generated by translating the dataset from English to German via DeepL, demonstrating that this is an acceptable approach for constructing a reference dataset for the German language. To

[1] DeepL SE, https://www.deepl.com/.

our knowledge, there is no comparable data set for the German language that identifies the use of hedging in scientific texts.

Analysis corpora

In our main analysis we differentiate between original literature, ASL and textbooks (Table 3). We sampled adapted literature written by ourselves in German ($n = 7$) and published in English by Yarden et. al (2015) ($n = 6$). The corpus of original literature consists of the corresponding English originals. Based on the language of the adaptation, we have assigned a mutual code to the corresponding text pair (EN1 to EN6 and DE1 to DE7). For comparison we randomly sampled eleven texts extracts from German biology and chemistry textbooks aiming 5th to 13th grade. A textbook text is always the narrative-explanatory part of a textbook chapter, which usually covers one or two pages. Assignments, captions, experiment instructions etc. have been removed from all textbook texts.

Table 3 Analysis corpora. A detailed list of the individual items in the corpus is provided in Table 7 to Table 9 (Appendix/SI). The number of sentences represents the arithmetic mean and one standard deviation.

Genre	**Language**	***n***	**Sentences**
Original literature	English	13	163 ± 79
Adapted scientific literature	English	6	136 ± 49
	German	7	122 ± 16
Textbook	German	11	34 ± 12

Preprocessing

Before the classification can be carried out, several steps must be completed (see Fig. 2). We scripted an XML parser which extracts the hedging cues from the XML files of the BioScope corpus and generates a CSV file, in which every sentence is labeled with *Hedging* or *No Hedging* depending on the presence of speculation cues. The resulting CSV file was then translated via DeepL from English to German.

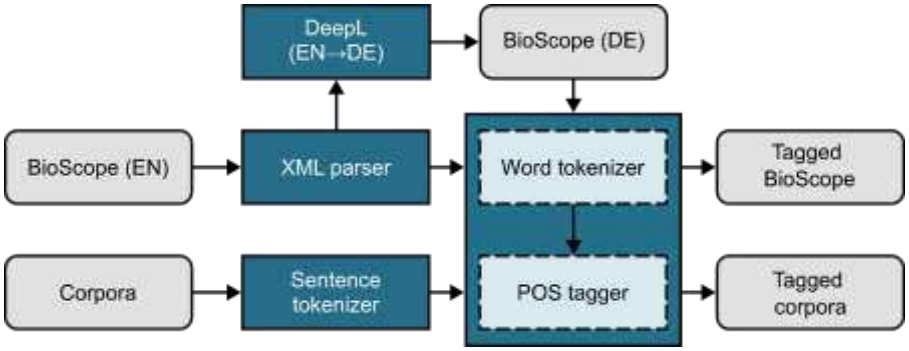


Fig. 2 Preprocessing pipeline

Since the analysis corpora were not pre-structured like the BioScope corpus, we had to tokenize the texts. Therefore we used the *PunktSentenceTokenizer* from the *Natural Language Toolkit* (NLTK; Bird et al., 2009) to split the texts in their sentences. We manually added some abbreviation to the routine of the tokenizer, which are common for scientific texts and textbooks: *ca.* (from "circa"), *approx.* ("approximately"), *Fig.* ("Figure") and respectively the German *Abb*. ("Abbildung"), *Tab.* ("table" or German "Tabelle"), *Ref.* ("reference"), *al.* (from "et al."), *etc*. and additionally for German *z. B.* (short form of "zum Beispiel" meaning "for example") and *usw*. ("und so weiter", alternative and more informal formulation for *etc*.).

In a subsequent step, all sentences were further tokenized into words, which were then tagged with their part of speech (POS) tag. We used the *TreebankWordTokenizer* from the NLTK and the *TreeTagger* from Schmid (1994, 1995) (with the help of the *treetaggerwrapper* from Pointal, 2019) with the BNC-Tagset for English (BNC Consortium, 2007) and the STTS-Tagset for German (Schiller et al., 1999) to perform this task. As a result, each text now consists of a list of sentences, which themselves are a list of tuples consisting of a word, its POS tag and the lemma. A lemma can be described as the dictionary form of a word or a set of words (e.g., the word *wrote* becomes its lemma *write*, while *me* → *I*), while the POS tag gives information about the grammatical function or word class of the word in the given context (e.g., superlative adjectives are coded as *AJS*). This simplifies the task immensely since we don't have to check every grammatical form of a word when searching. Table 4 gives an example to illustrate the process.

Table 4 Sample result of the tagging process. The table illustrates the tagging process using the first three sentences of the article by Crick and Watson (1954) on the double helix of DNA. *SENT* represents the end-of-sentence part-of-speech tag, while the meaning of the other POS tags can be found in the respective handbook (BNC Consortium, 2007).

Sentence	**Tagged version** (word, POS tag, lemma)

1	(This, DT0, this), (paper, NN1, paper), (describes, VVZ, describe), (a, AT0, a), (possible, AJ0, possible), (structure, NN1, structure), (for, PRP, for), (the, AT0, the), (paracrystalline, AJ0, paracrystalline), (form, NN1, form), (of, PRF, of), (the, AT0, the), (sodium, NN1, sodium), (salt, NN1, salt), (of, PRF, of), (deoxyribonucleic, AJ0, deoxyribonucleic), (acid, NN1, acid), (., SENT, .)
2	(The, AT0, the), (structure, NN1, structure), (consists, VVZ, consist), (of, PRF, of), (two, CRD, two), (DNA, NN1, dna), (chains, NN2, chain), (wound, VVD, wind), (helically, AV0, helically), (round, PRP, round), (a, AT0, a), (common, AJ0, common), (axis, NN1, axis), (,, PUN, ,), (and, CJC, and), (held, VVD, hold), (together, AV0, together), (by, PRP, by), (hydrogen, NN1, hydrogen), (bonds, NN2, bond), (between, PRP, between), (specific, AJ0, specific), (pairs, NN2, pair), (of, PRF, of), (bases, NN2, base), (., SENT, .)
3	(The, AT0, the), (assumptions, NN2, assumption), (made, VVN, make), (in, PRP, in), (deriving, VVG, derive), (the, AT0, the), (structure, NN1, structure), (are, VBB, be), (described, VVN, describe), (,, PUN, ,), (and, CJC, and), (co-ordinates, NN2, co-ordinate), (are, VBB, be), (given, VVN, give), (for, PRP, for), (the, AT0, the), (principal, AJ0, principal), (atoms, NN2, atom), (., SENT, .),

Training and classification

The tagged corpora form the basis for the classification process. In the next step, each sentence of the tagged BioScope and analysis corpora runs through the feature extractor. Based on the list of speculation keywords from the BioScope corpus, we have compiled 110 features for English and 142 features for German. We made sure that the words or phrases were good indicators of hedging, cleaned up any duplications or hedges that were too similar and transferred them to their lemma. For example, we removed *address a number of questions* as a feature, as *question* was already coded as a feature. The list of English features was then translated by us, which explains the increase in the number of German features. A list of all features is available at our code repository[2].

All features are binary coded: *True*, if the respective feature appears in the sentence and *False*, if not. By default, our feature extractor (Fig. 3) is checking for the presence of the lemma of a feature. However, this is only possible if the feature consists of a lemma that cannot occur in different grammatical forms that are to be included or

[2] See section code availability.

excluded. In German, this constitutes a problem because compounds usually occur in closed form and prefixes are separable from their verb but important for the semantic. For example, the verb *[etwas] ableiten* (meaning *derive [something]*) can be separated to *leiten [etwas] ab*, but the verb *leiten* without the prefix *ab* would mean something different (namely *guiding*). Fortunately, the STTS tagset recognizes these prefixes and marks them accordingly. Therefore, for each feature we developed alternative writings which were also checked. In terms of the previous example, this means that the feature *ableiten* is true if (a) the lemma *ableiten* is present or (b) if the lemma *leiten* as well as the lemma *ab* are present, with the important condition that *ab* has to have the corresponding POS tag of a separated particle of a verb (*abgetrennter Verbzusatz*, PKTVZ). In addition, for hedges in which negations occur, not every possible negation particle is checked, but only for a general negation POS tag. Also, German auxiliary verbs a tested differently: Instead of the lemma, the original word will be checked, since the lemma would be too vague.

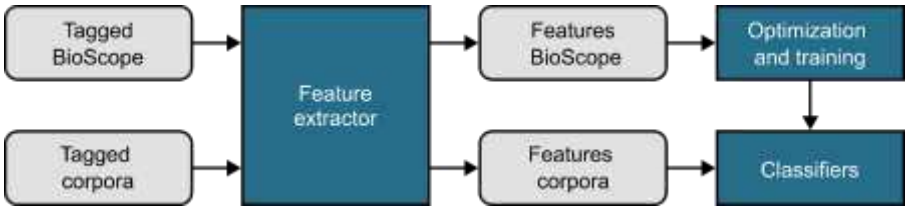


Fig. 3 Training and classification pipeline

As models we used a naive bayes classifier (NBC) from the NLTK (Bird et al., 2009) as well as a linear support vector classifier (SVC) from the *scikit-learn* package (Pedregosa et al., 2011) for each language. In a 5-fold cross validation we trained and optimized all four models. For the linear SVCs, we determined the best candidate for the parameter C between 10^{-3} and 10^{3} (see Fig. 8, Appendix). Since the BioScope corpus consists of more sentences without than with hedging, we corrected C with the respective class weight ω_i, which is inversely proportional to the class frequency[3]. Table 5 lists quality metrics of the cross validation.

Table 5 True positive rate, true negative rate, positive predictive value, accuracy, and F_1 score of the NBCs and SVCs (mean and SD of the 5-fold cross-validation). C values of $C = 10^{-0,5}\ \omega_i$ (SVC-EN) and $C = \omega_i$ (SVC-DE) were selected for the SVCs. The formulas for calculating the metrics are listed in the appendix/SI.

Metric	**NBC**		**SVC**	
	English	**German**	**English**	**German**

[3] $\omega_{\text{hedging}} \approx 2.8$, while $\omega_{\text{no hedging}} \approx 0.6$.

True positive rate	0.85 ± 0.02	0.73 ± 0.02	0.94 ± 0.01	0.83 ± 0.02
True negative rate	0.985 ± 0.001	0.988 ± 0.004	0.960 ± 0.004	0.964 ± 0.004
Positive predictive value	0.926 ± 0.008	0.93 ± 0.02	0.84 ± 0.01	0.84 ± 0.02
Accuracy	0.962 ± 0.005	0.952 ± 0.006	0.956 ± 0.004	0.940 ± 0.006
F_1 score	0.89 ± 0.02	0.82 ± 0.03	0.88 ± 0.01	0.83 ± 0.02

The quality of the classifiers can be summarized as good. It is noticeable that the classifiers trained on German texts show a poorer classification performance and classify more texts falsely negative compared to the English classifiers. Compared to the NBCs, the SVCs show a higher rate of true positive results with a poorer positive predictive value. As expected, the accuracy is quite high in all cases, as 82% of the corpus consists of sentences without hedging.

Analysis

All statistical operation were performed using R (R Core Team, 2023). For testing and plotting we used the *rstaticx* (Kassambara, 2023) and *ggplot2* package (Wickham, 2016) as well as *gnuplot*.

To test whether we can compare texts originating from different languages, we first investigated (a) to what extent the classification results of different languages are comparable and (b) whether the translation of DeepL gives similar results to an authentic German version of an article, before analyzing the main corpus. For this purpose, we classified six articles from the *Angewandte Chemie* journal and the corresponding *Angewandte Chemie International Edition* journal (Table 10, Appendix) and compared their results, demonstrating that via the translation no essential hedging features are altered. Based on these results, we decided to proceed with the analysis of the main analysis corpus and we also pooled the adapted scientific articles of English and German language into one group. A full description of the results is available in the appendix.

Results and discussion

The analysis of the results of the sentence classification explores three main ideas: (1) the overall comparison between the genres, i.e., authentic and adapted literature and textbook, (2) the distribution of hedging within the text and (3) the direct comparison of adaptation with its original.

In general, both classifier models, NBCs and SVCs, show similar results. On average, both models agree to 95 ± 4 %; This is fairly constant across all three text genres. On a first glance, the range of results in the different text groups is quite large and between 0 to 39.5 % (NBC) and 0 to 51.3 % (SVC) sentences with hedging.

Comparison of genres

On average, 2 ± 3 % (NBC) respectively 5 ± 4 % (SVC) of the sentences in a textbook text use uncertainty expressions, whereof five (NBC) respectively three (SVC) texts don't use hedging at all. The original literature shows a high usage of hedging: 20 ± 9 % (NBC) respectively 26 ± 11 % (SVC) of the sentences are classified as hedging. Slightly below, but still considerably more compared to the textbooks, are the adaptations with 13 ± 7 % (NBC) respectively 18 ± 9 % (SVC). Fig. 4 visualizes the results as boxplot.

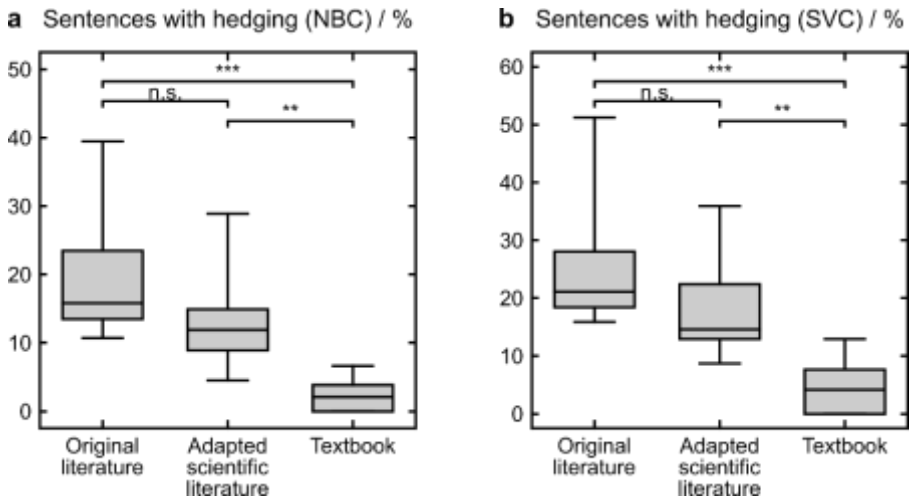


Fig. 4 Boxplots depicting the proportion of hedging in the analysis corpus by genre. (a) Results of the NBCs and (b) the SVCs. Significant differences can be found between the gernes textbooks and adaptations as well as textbooks and original literature (Kruskal-Wallis test and post hoc Dunn test with Bonferroni adjustment; **: $p < 0.01$, ***: $p < 0.001$, n.s.: no significance).

A Kruskal-Wallis Test was conducted to examine the difference between the gernes. Significant differences were found for both classifier models (NBC: $\chi^2(2) = 23.991,\ p < 0.001$; SVC: $\chi^2(2) = 23.578,\ p < 0.001$). Subsequent post-hoc tests (Dunn-Bonferroni tests) show significant difference between original literature and textbooks (NBC: $z = -4.83,\ p < 0.001, r = 1$; SVC: $z = -4.80, p < 0.001, r = 1$) as well as adaptations and textbooks (NBC: $z = -3.29, p = 0.003, r = 0.7$; SVC: $z = -3.23, p = 0.004, r = 0.7$), while adaptations and original literature show no significant differences (NBC: $z = 1.61,\ p = 0.3$; SVC: $z = 1.64,\ p = 0.3$). These results confirm the expectation of a reliable text style in textbook texts. Furthermore, it seems like the authenticity of original scientific literature in terms of hedging is not comprised after the adaptation.

Similar to the results of the *Angewandte Chemie* corpus (see Appendix), the distribution of the results is quite wide. It is possible that many different effects play a role, e.g., the authorship of the text and related to that the cultural background of the author(s), the topic, the publication date or the journal. To our knowledge there are little

to no studies exploring the impact of these properties on the hedging of an article. A preprint article from Costello et al. (2023) suggests a correlation between the frequency of hedging in an article and the gender of their authors as well as demographics of the assigned editor. This indicates a complex relation between hedging and different parameter. Nevertheless, our results show that the common text gerne for secondary education uses little to no hedging and therefore depicts science as rather certain, while original and adapted scientific literature use hedging in nearly every tenth to fifth sentence.

Comparison between original and adaptation

When focusing on original literature and adaptation, we shift the analysis from a genre to a text level and compare adaptations with their corresponding original. This allows a clear view on the effect of adaptation on the usage of hedges. On average, the number of sentences classified with hedging is reduced by 15 ± 20 (NBC) respectively 19 ± 27 (SVC) sentences compared to the original literature. In three cases (NBC: DE7; NBC and SVC: EN3, EN6) more sentences in the adapted article are classified with hedging. Since the total number of sentences varies heavily and changes between the genres, we compare the proportion of sentences with hedging. With the exception of two texts (SVC: EN6; NBC and SVC: EN3), the proportion of sentences of the adaptations are lower than compared to their originals (see Fig. 5).

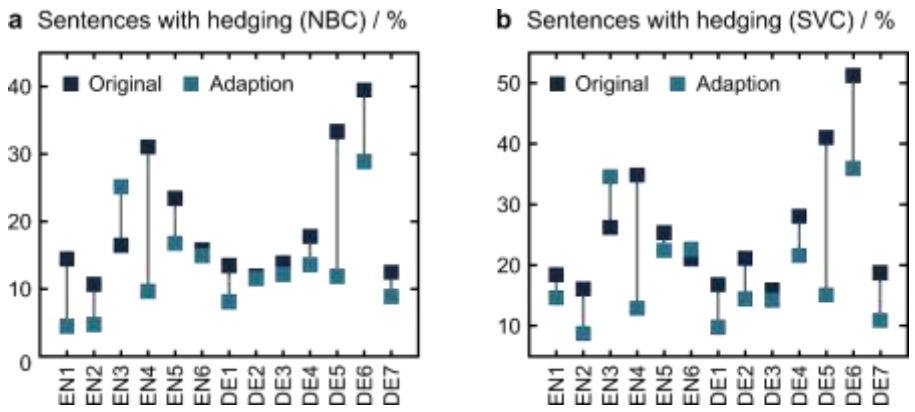


Fig. 5 Pairwise comparison between original text and adaptation by classifier (a: NBC, b: SVC)

By dividing the proportion of sentences with hedging of the adapted article with these of the original article, we calculated the factor of preservation of hedging for both language combinations (see Table 6). In total, the proportion of hedging is reduced by a factor 0.7 ± 0.3 (NBCs as well as SVCs). However, the factor lies within a broad range of 0.3 to 1.5. Interestingly, the changes in the German adaptations are more homogeneous: Except for DE5, the factor of all articles are closely, as also indicated by the smaller standard deviation in Table 6. This is different for the English language, where a broader range can be observed.

Table 6 Preservation of hedging compared between original text and adaptation by classifier and language combination (mean and one standard deviation)

Language combination	***n***	**Factor of preservation of hedging**	
		NBC	**SVC**
English-English	6	0.7 ± 0.5	0.8 ± 0.4
English-German	7	0.7 ± 0.2	0.7 ± 0.2
Total	13	0.7 ± 0.3	0.7 ± 0.3

The heterogeneity of the preservation provokes the question if the adaptation process is reliable. A systematic pattern in these results is initially not recognizable. It can be assumed that the respective author of the adapted article has a decisive influence on the use of hedging. This could explain the lower variance in the German articles, which were produced in the research group of the authors, while the authors of the examined English articles are much more divers. Especially an influence of the language of the article as well as the cultural background of the authors could have an impact on the use of hedges.

When comparing each original with its respective adapted article (Fig. 5), we want to discuss the results of the combinations EN3, EN4, and DE5 separately, since they show a prominent loss or gain of sentences with hedging. The articles EN3 discuss the phenomenon that the solar corona is hotter than lower layers of the sun, also known as the "coronal heating problem" (Aschwanden et al., 2007; Yarden et al., 2015). While the original discusses ten different arguments for a solution of this paradox, the author of the adapted article focused on only three arguments, but the length of both articles is nearly the same (about 160 sentences). This means, the authors added additional content in the article, possibly for making the arguments more accessible. According to Yarden et al., the adapted paper "demonstrates a genuinely controversial issue" and "illustrates how scientists go about trying to resolve controversy" (2015, p. 179). Since the article doesn't present new data but rather discusses evidence, the additional explanations could also use hedges to express uncertainty, which could explain the increase of hedging.

In contrast to EN3, the article pair EN4 shows a significant loss of sentences with hedges in the adapted article (factor of preservation = 0.3, NBC, respectively 0.4, SVC). With only 94 sentences, the adapted article is one of the shortest of the articles examined. In addition, following the descriptions of Yarden et al. (2015), the introduction and discussion were intensively supplemented with explanations so that the text is understandable for the target group of the article. Taken both together, both factors could at least partially explain the larger relative decrease in hedging.

Finally, similar to EN4, the article pair DE5 shows an extreme decrease of the hedging density from the original to the adapted article (factor of preservation = 0.4). The original article is a review article from Bouteldja and Timson (2010) exploring the biochemical basis of hereditary fructose intolerance, focusing especially on the A149P mutation of the aldolase B protein. It is possible that this could indicate structural differences between article types, since EN4 and DE5 could be described as review articles. On the contrary, also the articles DE4 are review article and do not show such a decrease. It would be interesting to analyze and compare the use of hedging controlled by the article type described by Harmon and Gross (2004).

The heterogeneity between the genres shows, that the use of hedging is still a matter of writing style. Not only the authors of the articles but also the authors of the adapted articles have to decide if they want to phrase a statement in an uncertain way, possibly weaking their arguments. A study from Horn (2001) investigated this phenomenon and his findings suggests that scientists, paraphrasing another author's findings and ideas, often reduce the uncertainty of the original statement, making them more certain. By comparing 50 quotes from five original statements from mainly biological articles, Horn calculated that the hedging is preserved in only 60 to 68% of all cases – which is, interestingly, in a similar quantity to our observation from original to adapted article (Table 6).

Distribution throughout the text

Similar to the study by Atanassova et al. (2018) we subsequently analyzed the distribution of hedging within the articles. Since the density of hedging changes within the article, as research from Atanassova et al. or Hyland (1996) shows, analyzing the distribution of hedging within the articles can not only answer the question, if ASL shows a similar distribution, which would be an indication for an authentic adaptation, but also whether the observed reduction in the hedging quantity occurs evenly over the entire text or to the disfavor of certain text sections. Therefore, we collected the position of a sentence which was classified with hedging for each genre and normalized it to a relative position of 0 to 100% in term of text progression. Fig. 6 describes the distributions of sentences with hedging by text genre (a: NBC; b: SVC) as well as the density distribution (Fig. c).

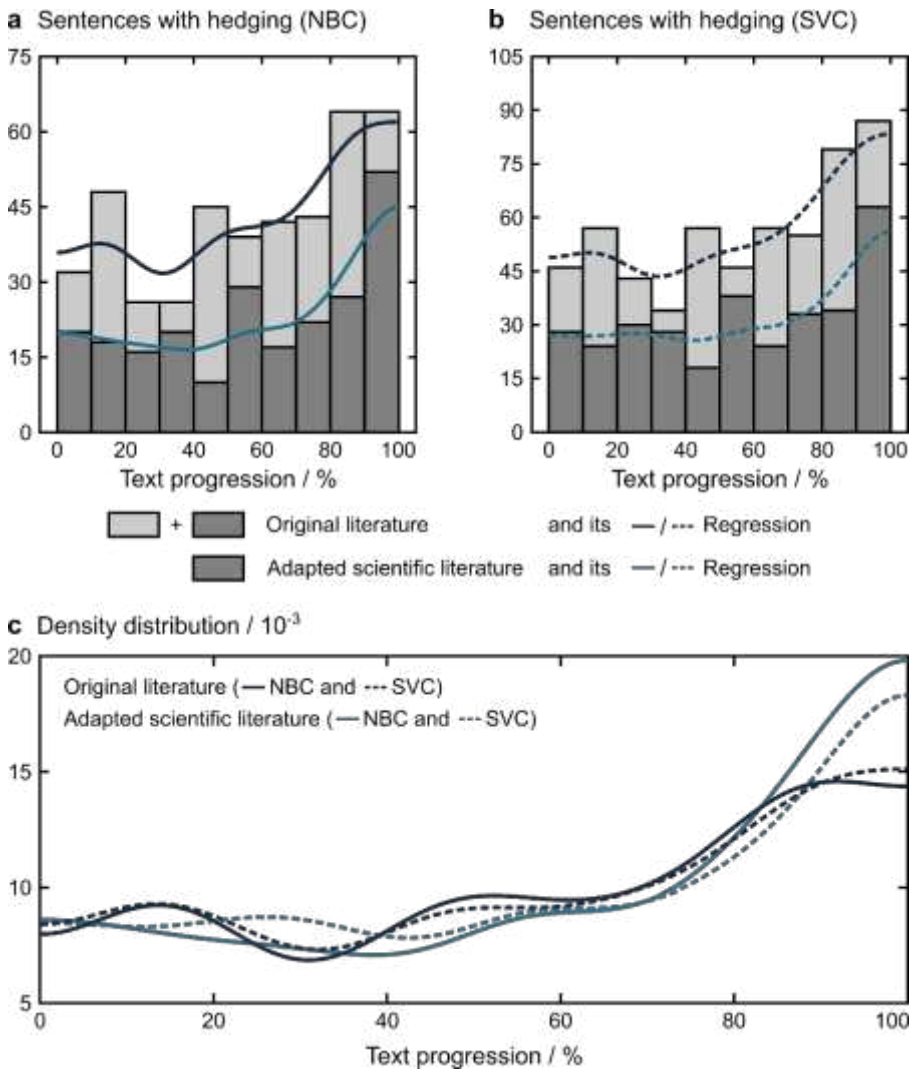


Fig. 6 Distribution of hedging over the body of the text per classifier and text genre. Fig. a and b show histograms of the sentences with hedging across all texts of the corresponding genre (dark gray: adaptation; dark gray and light gray aggregated: original literature; class width 10% in each case for comparability). The superimposed curves are the result of a kernel regression (gaussian kernel smoother, bandwidth 10%, limited from 0 to 100 %). Fig. c shows the density of the distribution by text genre and classifier, each normalized to an integral of one (kernel regression with gaussian kernel smoother, separate calculation of the bandwidth according to Silverman, 1986, limited from 0 to 100% in each case).

Overall, the distribution of hedging is quite similar between the genres. Nevertheless, the reduced use of hedging in the adaptations compared to the original can also be observed here, tough the differences do not spread evenly. For both genres and both classifier models, a significant increase in frequency can be observed in the last third of the text progression, where the discussion section is located according to the classic IMRaD model (Introduction, Methods, Results and Discussion). Since the goal of the discussion is to contextualize the presented results, weigh

uncertainties and biases and argue about the reliability of the interpretation of these results, a higher density of sentences requiring a moderation of their semantic strength is expected. Results from Atanassova et al. (2018) and Hyland (1996) confirm this expectation empirically. Following this, the examined articles, both originals and adapted versions, are authentic.

Differences are observable between 10 to 40 % and 90 to 100 % (see Fig. 6c). It is debatable whether this is caused by the small sample size or if the differences are significant. Assuming the latter, the increase of hedging density between 10 and 20 % as well as the decrease in hedging density between 20 to 40 % text progression of the original article could be explained with the appearance of the method section, which describes the used methods and materials and therefore does not need hedging. In contrast, the introduction (usually at the beginning of an article) describes preliminary results and discusses theories – often accompanied with uncertainties (Atanassova et al., 2018; Hill et al., 1982; Hyland, 1996). We hypothesize that the adaptation of the introduction leads to a lower hedging density in the introduction in an adapted article, similar to the methods section, resulting in a relatively stable density for the first 40 % of the text progression (compare with Fig. 6). The general aim of an introduction is it to lead the reader from the general topic to the specific context of the article. Because the adapted article needs to respect a different prior knowledge of their reader, i.e. students, more concepts have to be explained, which are considered as relatively certain by the scientific community. This could explain the lesser use of hedges in the first part of adapted articles compared to the original articles.

Our results suggest a relatively high density at the end of adapted articles compared to the original articles (between 90 to 100 % text progression). As shown in Fig. 6a and b, the differences of the quantity of sentences with hedging between adapted and original articles is smaller in the 20 to 40 % and 90 to 100 % intervals, meaning the preservation of hedging throughout the adaptation process is higher in these areas. Since the preservation of hedging is normally lower (Table 6), this results in a relatively high density for the adapted articles.

Limitations

The results of this analysis depend directly on three key points: the quality of the tokenizers, the quality of the reference corpus as well as the quality of the analysis corpus. If the sentence tokenizer overtokenizes, the text will be split into too many sentences, resulting in an underestimation of the proportion of sentences with hedging. This could be to the disadvantage of languages or text genres which frequently use abbreviations with the use of periods that are unknown to the tokenizer.

Directly related to this is the quality of the training data and the selection of features. Since the BioScope dataset is mainly based on abstracts of biomedical articles and the features are extracted from them, there could be a bias in the classification process. It is possible that the use of hedging depends not only on the authors but also on the scientific culture of the scientific domain and therefore we systematically under detect hedging in non-biomedical articles. Unfortunately, we are not aware of scientific research comparing the use of hedging between different scientific domains. Nevertheless, the cross-validation showed good results on the generalization for unknown data, so we believe that this effect is rather small.

Additionally, we had to rely on the translation quality of DeepL when generating training data for the classification of German articles. Although our analysis on bilingual articles from *Angewandte Chemie* showed a 90 to 92 % agreement between both languages (Table 11, Appendix), we still compare two different languages. This is, like the authors or the different scientific domains, another element which is expected to have individual influences and necessitaes further scientific exploration.

For the analysis of the distribution throughout the text progression, we had to assume a similar structure within the articles. Each article is at least three-parted (Introduction, Procedure, Discussion), and even most article we examined follow the IMRaD structure. However, the length of the articles varies, with the shortest article consisting of 64 sentences and the longest of 334 sentences. In our interpretation we had to assume a similar relative structure, i.e. a comparable length of each section, in order to attribute the density of hedging to certain parts of a typical scientific article.

Conclusion

Despite the limitations described above and the relatively small sample size, we showed that although adapted articles have a lower overall proportion of sentences with hedging than their originals, their distribution within the text is similar to the originals. However, as with the originals themselves, the use of hedging presumably depends on many factors, above all the authorship of the text. The dispersion of hedging between articles from the *Angewandte Chemie* shows that there are also differences within a journal, which can, generally speaking, also apply on the authors of original and adapted articles.

In the introduction we raised the question as to whether adapted literature is authentic scientific literature, i.e. whether the adaptation preserves important, characteristic features of the original. Using the example of hedging, this question can be answered in the affirmative, considering the limitations described. In addition to the results of Ariely et al. (2019), describing a functional correspondence in the use of hedging between original and adaptation,

we now can report on a positional as well. Only the quantity of hedging is reduced, but it often reaches at least half of the original and 70 % on average. For a complete answer on the authenticity of ASL, however, other linguistic and structural features would have to be examined and compared in addition to hedging, such as the use of passive voice, personal pronouns, citations or graphics and tables. However, this study emphasizes the usability of ASL in educational settings to convey authentic features of science and scientific literature by a new genre, which is easier to understand for novice, but stay authentic in its scientific features. As proposed by other authors (I. Braun & Nückles, 2014; Hagen et al., 2022; Phillips & Norris, 2009; Yarden et al., 2001), this could open new avenues for fostering a deepened understanding of the features of science.

Appendix

Corpora

```
<Annotation created="23/6/2008" creator="XMLconverter">
  <DocumentSet>
    <Document type="Biological_full_article">
      <DocID type="BMC_ID">1471-2105-8-225</DocID>
      ...
      <DocumentPart type="SubSectionTitle">
        <sentence id="S1.3">Background</sentence>
      </DocumentPart>
      <DocumentPart type="Text">
        <sentence id="S1.4">Selenocysteine and pyrrolysine are
            the 21st and 22nd amino acids, which are genetically
            encoded by stop codons.</sentence>
        <sentence id="S1.5">Since a number of microbial genomes
            have been completely sequenced to date, it is
            tempting to ask <xcope id="X1.5.1"><cue
            type="speculation" ref="X1.5.1">whether</cue> the
            23rd amino acid is left undiscovered in these
            genomes</xcope>.</sentence>
        ...
      </DocumentPart>
      ...
    </Document>
    ...
    </DocumentSet>
</Annotation>
```

Fig. 7 Excerpt from the BioScope XML file containing the full articles (Vincze et al., 2008)

Table 7 Examined adaptations and their originals (in English)

Code	Adapted article	Original article
EN1	Developing an Inhibitor of Anthrax Toxin (Yarden et al., 2015, pp. 146–154)	Designing a polyvalent inhibitor of anthrax toxin (Mourez et al., 2001)
EN2	Expression of the *Bacillus thuringiensis* (*Bt*) Toxin in Chloroplasts of Tobacco Plants Imparts Resistance to Insects (Yarden et al., 2015, pp. 223–234)	Overexpression of the *Bt cry2*Aa2 operon in chloroplasts leads to formation of insecticidal crystals (Cosa et al., 2001)
EN3	The Coronal Heating Problem (Yarden et al., 2015, pp. 180–195)	The Coronal Heating Paradox (Aschwanden et al., 2007)
EN4	Absence of Skeletal Muscle in Mouse Embryos Carrying a Mutation in the *Myogenin* Gene (Yarden et al., 2015, pp. 215–222)	Muscle deficiency and neonatal death in mice with a targeted mutation in the *myogenin* gene (Hasty et al., 1993)
EN5	Did Early Maritime Archaic Indians Ever Live in Newfoundland? (Yarden et al., 2015, pp. 198–207)	Prehistoric cultures, reconstructed coasts (Bell & Renouf, 2004)
EN6	West Nile Virus: When Will an Outbreak Occur and How Can We Prevent It? (Yarden et al., 2015, pp. 162–178)	An epidemiological model for West Nile Virus (Wonham et al., 2004)

Table 8 Examined adaptations and their originals (German adaptations and English originals).

Code	Adapted article	Original article
DE1	Spannungsabhängige Konformationsänderungen des Ionenkanals VDAC-1: Ein Vergleich zwischen Simulation und Experiment (von Hoff et al., 2019)	Voltage Dependence of Conformational Dynamics and Subconducting States of VDAC-1 (Briones et al., 2016)
DE2	Synthese von Glycosphingolipiden mit fluoreszenzmarkierten Fettsäuren für die quantitative Untersuchung der Anbindung von Shiga-Toxin an Lipidmembranen (von Hoff, 2020)	Gb_3 Glycosphingolipids with Fluorescent Oligoene Fatty Acids (Patalag et al., 2017)
DE3	SNARE-vermittelte Membranfusion an festkörpergestützten und freitragenden Lipidmembranen (not published)	SNARE-Mediated Single-Vesicle Fusion Events with Supported and Freestanding Lipid Membranes (Kuhlmann et al., 2017)
DE4	Toxizität und Kanzerogenität von Glyphosat: Ein Review über die wissenschaftlichen Grundlagen der Untersuchung der Europäischen Union und dessen Unterschiede zum IARC (Aljets, 2024)	Glyphosate toxicity and carcinogenicity: a review of the scientific basis of the European Union assessment and its differences with IARC (Tarazona et al., 2017)
DE5	Die biochemische Basis vererblicher Fructose-Intoleranz (not published)	The biochemical basis of hereditary fructose intolerance (Bouteldja & Timson, 2010)
DE6	Ein Vorschlag zur biochemischen Struktur der Desoxyribonukleinsäure (DNS) (not published)	The complementary structure of deoxyribonucleic acid (Crick & Watson, 1954)
DE7	Die photochemisch angetriebene umgekehrte Wassergas-Verschiebung – ein attraktiver Ansatz für eine klimafreundliche Herstellung von Synthesegas? (not published)	Photochemically Driven Reverse Water-Gas Shift at Ambient Conditions mediated by a Nickel Pincer Complex (Schneck et al., 2018)

Table 9 Examined text sections from biology and chemistry textbooks (German). T11 to T14 were randomly selected from lower secondary school textbooks, while T21 to T27 specifically represent texts on the molecular structure of DNA.

Code	Title of text section	Textbook
T11	Das Teilchenmodell und der Aufbau der Stoffe	Elemente Chemie 7-10 Niedersachsen G8 (Irmer et al., 2010, pp. 56–57)
T12	Was Atome in Molekülen zusammenhält	Chemie heute SI Gesamtband (Asselborn et al., 2013, pp. 194–195)
T13	Die Atomhülle	Fokus Chemie SI Gesamtband (Jaek et al., 2017, pp. 98–99)
T14	Mach mal Salz! – Halogene	NEO Chemie SI Niedersachsen Gesamtband (Kraft et al., 2015, pp. 134–135)
T21	Nucleinsäure - DNA und RNA	Chemie im Kontext (Demuth et al., 2006, pp. 350–351)
T22	DNA - der Stoff aus dem die Gene sind	Chemie heute SII Gesamtband (Asselborn et al., 2009, pp. 458–459)
T23	Erbsubstanz DNA	Linder Biologie (Bayrhuber et al., 2010, pp. 95–96)
T24	Die Nucleinsäuren DNA und RNA	Natura – Oberstufe, Biologie für Gymnasien (Baak et al., 2016, pp. 146–147)
T25	Im DNA-Molekül bilden zwei Nucleotidstränge eine Doppelhelix	Markl Biologie Oberstufe (Markl, 2018, pp. 163–165)
T26	Molekularer Aufbau von DNA und RNA	Biologie heute SII (J. Braun et al., 2011, pp. 121–122)
T27	Die DNA – ein sehr großes Molekül	Erlebnis Naturwissenschaften 3 (Cieplik et al., 2015, pp. 55–56)

Cross-validation and metrics

The number of true positive (TP), false positive (FP), true negative (TN) and false negative (FN) cases was determined and used for the calculation of the quality metrics true positive rate / recall (TPR), true negative rate / specificity (TNR), positive predictive value / precision (PPV), accuracy (ACC) and F_1 score:

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}$$

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{N}$$

$$F_1 = \frac{2\,\mathrm{TP}}{2\,\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}$$

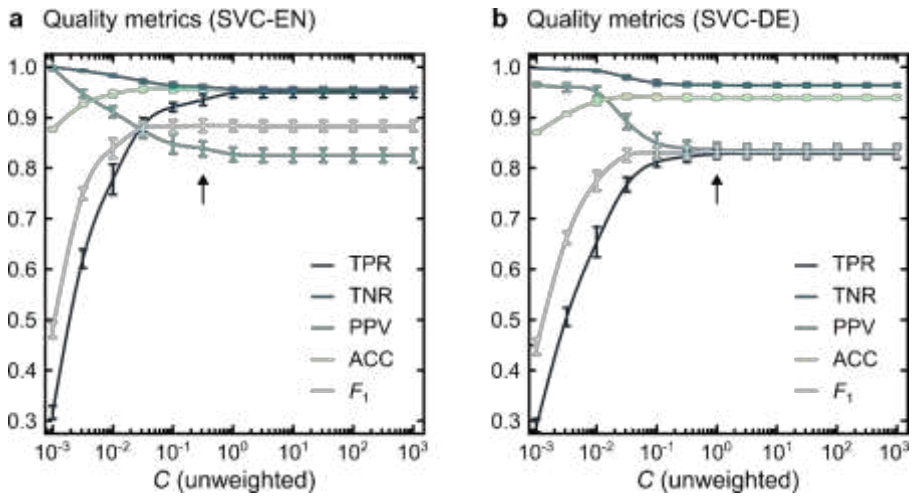


Fig. 8 Sensitivity, specificity, precision, accuracy and F1 score of the linear SVC as a function of the parameter *C*. (a) Results for the linear SVC of English language (SVC-EN) and (b) German language (SVC-DE). The arrows indicate the results of the chosen *C* values. Please note that the *C* value will be corrected by the class weight ω_i.

Comparability of the classification of English and German language

We gathered twelve articles from the *Angewandte Chemie* and the corresponding *Angewandte Chemie International Edition* journals (Table 10). Since *Angewandte Chemie* is a traditional German journal, authors of articles have the opportunity to publish their article in German as well as in English. This gives us the opportunity to compare scientific texts of English and German language in an authentic way as both versions are genuine and not created by a translation software. Additionally, we translated the articles from *Angewandte Chemie International Edition* from English to German using DeepL, so the corpus consists of six articles (AC1 to AC6) in three different "languages": English (EN), German (DE) and a translation from English to German (EN→DE).

To ensure a comparability between the three different article versions of the *Angewandte Chemie* corpus, we checked the results of the sentence tokenization and manually corrected the articles to the same number of sentences. During the process, it was ensured that sentences that belong together in relation to their content are in the same position in all three variants, so that a direct comparison is possible. However, this also caused sentences to be merged, since some sentences were split into two in the other variant of the article. For reasons of simplicity, we will continue to speak of sentences.

Table 10 Examined *Angewandte Chemie* articles. All articles are Open Access titles from the May issue of *Angewandte Chemie* (Vol. 135, Issue 22) or *Angewandte Chemie International Edition* (Vol. 62, Issue 22). Since the journal published bilingual, an English and a German version of the article is available for analysis.

Code	English	German
AC1	Close-Shell Reductive Elimination versus Open-Shell Radical Coupling for Site-Selective Ruthenium-Catalyzed C−H Activations by Computation and Experiments (X. Chen et al., 2023a)	Reduktive Eliminierung in geschlossener Schale und radikalische Kopplung in offener Schale für regioselektive Ruthenium-katalysierte C−H-Aktivierungen anhand von Berechnungen und Experimenten (X. Chen et al., 2023b)
AC2	Multimodal Analysis of Light-Driven Water Oxidation in Nanoporous Block Copolymer Membranes** (Kund et al., 2023b)	Multimodale Analyse lichtinduzierter Wasseroxidation in nanoporösen Blockcopolymer-Membranen** (Kund et al., 2023a)
AC3	Reduction of Platinum(IV) Prodrug Hemoglobin Nanoparticles with Deeply Penetrating Ultrasound Radiation for Tumor-Targeted Therapeutically Enhanced Anticancer Therapy (Liang et al., 2023a)	Reduktion von Platin(IV)-Prodrug Hämoglobin Nanopartikeln mit tief eindringender Ultraschallbestrahlung für eine gezielte und therapeutisch verstärkte Krebstherapie (Liang et al., 2023b)
AC4	Harnessing the Structure and Dynamics of the Squalene-Hopene Cyclase for (−)-Ambroxide Production** (Schneider et al., 2023a)	Verwendung der Struktur und Dynamik der Squalen-Hopen-Zyklase für die (−)-Ambroxid Produktion** (Schneider et al., 2023b)
AC5	Synthesis and Photophysical Properties of Monometallic C^C* Platinum(II) Formamidinate Complexes using Sterically Demanding Ligands (Wurl et al., 2023b)	Synthese und photophysikalische Eigenschaften monometallischer C^C* Platin(II)-Formamidinatkomplexe mit sterisch anspruchsvollen Liganden (Wurl et al., 2023a)
AC6	Photocatalytic C−C Cleavage of Methylenecyclobutanes for γ,δ-Unsaturated Aldehydes by Strain Release (Zhang et al., 2023a)	Photokatalytische Ringöffnung von Methylencyclobutanen zur Synthese von γ,δ -ungesättigten Aldehyden (Zhang et al., 2023b)

The results of the direct comparison for each sentence can be found in Table 11. A majority of the sentences were classified consistently across all languages (91 % for NBCs and 89 % for SVCs). Similar agreement is reached between the original English (EN) and German (DE) articles as well as the English and the translated articles (EN→DE). Comparing the classification of the sentences of the original German and the translated articles, the agreement is even higher, reaching 97 %.

Table 11 Sentence classification consensus between the different variants of the *Angewandte Chemie* corpus. Shown is the mean with a standard deviation across all six articles.

Classifier model	Consensus / %			
	All three	**EN and DE**	**EN and EN→DE**	**DE and EN→DE**
NBC	91 ± 5	92 ± 1	93 ± 5	97 ± 2
SVC	89 ± 2	90 ± 1	91 ± 2	97 ± 2

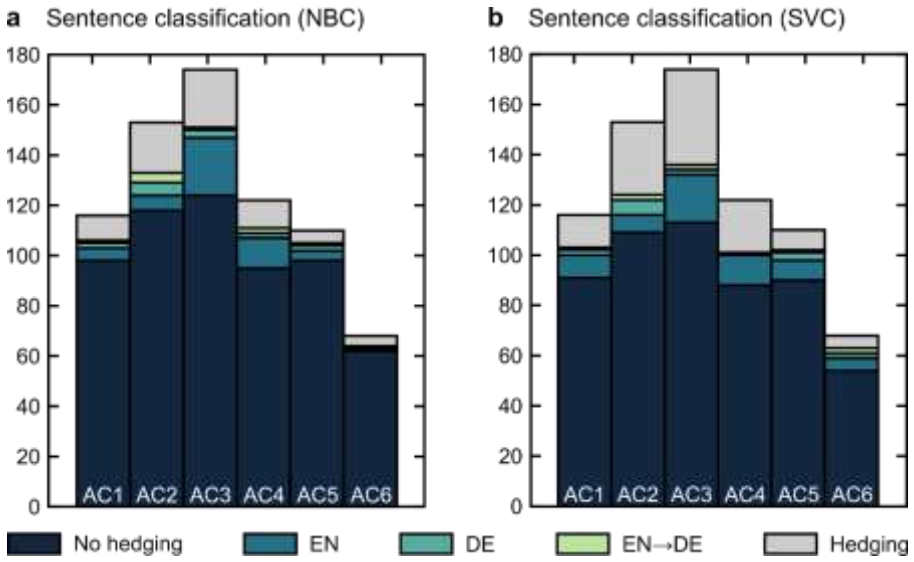


Fig. 9 Histogram of the sentence classification of the *Angewandte Chemie* corpus. Under the language-specific groups (EN, DE and EN→DE), sentences that were classified exclusively positive or negative in contrast to the other two variants are indicated.

Changing the focus from the single sentence to the whole article results in a better agreement between the languages, as the classification results are balanced with each other. Comparing the proportion of sentences classified with hedging, the differences reduce drastically. On average, the proportions of positively classified rates differ by 1.8 percentage points (NBC) and 3.1 percentage points (SVC) respectively (see Fig. 10 for the results of each article), with a maximum difference of 4.6 % (NBC) between the EN and EN→DE version of AC3 and 4.3 % (SVC) between the EN and DE version of AC1.

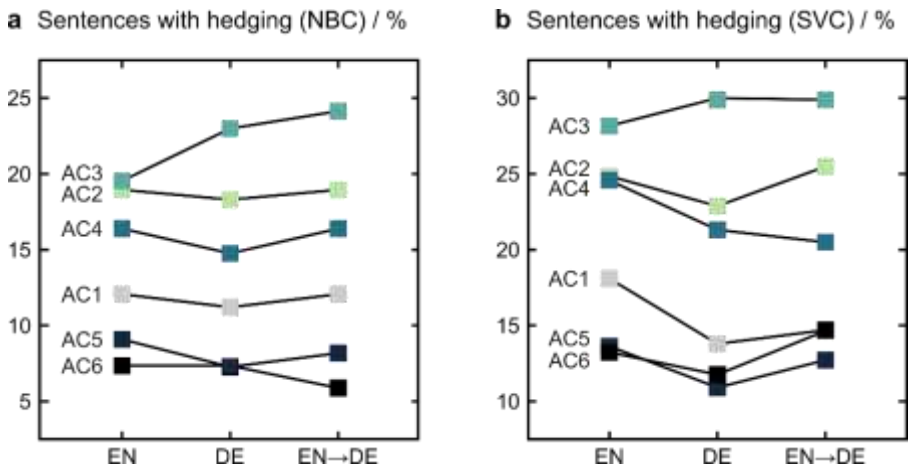


Fig. 10 Proportion of positively classified sentences from the English (EN), German (DE) and translated (EN→DE) *Angewandte Chemie* corpus. Fig. a shows the results of the NBCs, while fig. b shows those of the SVCs. Please note the restrictions to the term "sentence" discussed in the article.

Overall, these results lead us to the following conclusions: (a) The classifications between the different languages are comparable with each other, as the deviations are within a small range. Since the differences are in a similar magnitude as the positive predictive value of the models (see Table 5), they can be explained by the classifier model and not by differences between the languages. However, this only holds true if we assume that the authors and the editors of the journal preserve hedging when translating the article. (b) The high agreement between the classification of DE and EN→DE *Angewandte Chemie* articles indicates a sufficient quality of the translation. Therefore, we decided to proceed with the analysis of the main analysis corpus and we also pooled the adapted scientific articles of English and German language into one group.

References

Aljets, H. (2024). *Adaptierte wissenschaftliche Literatur. Charakteristika und Potential zur Vermittlung von Nature of Science* [PhD, Georg-August-Universität Göttingen]. https://doi.org/10.53846/goediss-10297

Ariely, M., Livnat, Z., & Yarden, A. (2019). Analyzing the Language of an Adapted Primary Literature Article. Towards a Disciplinary Approach of Science Teaching Using Texts. *Science & Education*, *28*(1–2), 63–85. https://doi.org/10.1007/s11191-019-00033-5

Aschwanden, M. J., Winebarger, A., Tsiklauri, D., & Peter, H. (2007). The Coronal Heating Paradox. *The Astrophysical Journal*, *659*(2), 1673. https://doi.org/10.1086/513070

Asselborn, W., Jäckel, M., & Risch, K. T. (Eds.). (2009). *Chemie heute SII Gesamtband*. Westermann Schroedel.

Asselborn, W., Jäckel, M., Risch, K. T., & Sieve, B. F. (Eds.). (2013). *Chemie heute SI Gesamtband*. Westermann Schroedel.

Atanassova, I., Rey, F.-C., & Bertin, M. (2018). Studying Uncertainty in Science: A distributional analysis through the IMRaD structure. In European Language Resources Association (Ed.), *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. http://lrec-conf.org/workshops/lrec2018/W24/summaries/10_W24.html

Baak, K., Becker, A., Eckebrecht, D., Kießling, J., Koch, M., Maier, A., & Roßnagel, G. (2016). *Natura—Oberstufe, Biologie für Gymnasien* (1st ed.). Ernst Klett Verlag.

Bayrhuber, H., Hauber, W., & Kull, Ulrich (Eds.). (2010). *Linder Biologie. Lehrbuch für die Oberstufe* (23rd ed.). Westermann Schroedel.

Bazerman, C. (1985). Physicists Reading Physics. *Written Communication*, *2*(1), 3–23. https://doi.org/10.1177/0741088385002001001

Bell, T., & Renouf, M. A. P. (2004). Prehistoric cultures, reconstructed coasts: Maritime Archaic Indian site distribution in Newfoundland. *World Archaeology*, *35*(3), 350–370. https://doi.org/10.1080/0043824042000185766

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python* (1st ed.). O'Reilly. https://www.nltk.org/book/

BNC Consortium. (2007). *British National Corpus, XML edition*. https://hdl.handle.net/20.500.12024/2554

Bouteldja, N., & Timson, D. J. (2010). The biochemical basis of hereditary fructose intolerance. *Journal of Inherited Metabolic Disease*, *33*(2), 105–112. https://doi.org/10.1007/s10545-010-9053-2

Braun, I., & Nückles, M. (2014). Scholarly Holds Lead Over Popular and Instructional: Text Type Influences Epistemological Reading Outcomes. *Science Education*, *98*(5), 867–904. https://doi.org/10.1002/sce.21117

Braun, J., Paul, A., & Westendorf-Bröring, E. (Eds.). (2011). *Biologie heute SII*. Schroedel.

Briones, R., Weichbrodt, C., Paltrinieri, L., Mey, I., Villinger, S., Giller, K., Lange, A., Zweckstetter, M., Griesinger, C., Becker, S., Steinem, C., & Groot, B. L. (2016). Voltage Dependence of Conformational Dynamics and Subconducting States of VDAC-1. *Biophysical Journal*, *111*(6), 1223–1234. https://doi.org/10.1016/j.bpj.2016.08.007

Chen, C., Song, M., & Heo, G. E. (2018). A scalable and adaptive method for finding semantically equivalent cue words of uncertainty. *Journal of Informetrics*, *12*(1), 158–180. https://doi.org/10.1016/j.joi.2017.12.004

Chen, X., Gülen, H. C., Wu, J., Zhang, Z.-J., Hong, X., & Ackermann, L. (2023a). Close-Shell Reductive Elimination versus Open-Shell Radical Coupling for Site-Selective Ruthenium-Catalyzed C−H Activations by Computation and Experiments. *Angewandte Chemie International Edition*, *62*(22), e202302021. https://doi.org/10.1002/anie.202302021

Chen, X., Gülen, H. C., Wu, J., Zhang, Z.-J., Hong, X., & Ackermann, L. (2023b). Reduktive Eliminierung in geschlossener Schale und radikalische Kopplung in offener Schale für regioselektive Ruthenium-katalysierte C−H-Aktivierungen anhand von Berechnungen und Experimenten. *Angewandte Chemie*, *135*(22), e202302021. https://doi.org/10.1002/ange.202302021

Cieplik, D., Freundner-Huneke, I., Kirks, H.-D., Tegen, H., & Zeeb, A. (Eds.). (2015). *Erlebnis Naturwissenschaften 3* (A). Westermann Schroedel.

Cosa, B. D., Moar, W., Lee, S.-B., Miller, M., & Daniell, H. (2001). Overexpression of the *Bt cry2*Aa2 operon in chloroplasts leads to formation of insecticidal crystals. *Nature Biotechnology*, *19*(1), Article 1. https://doi.org/10.1038/83559

Costello, A., Fedorova, E., Jin, Z., & Mihalcea, R. (2023). *Editing a Woman's Voice* (Preprint arXiv:2212.02581). arXiv. https://doi.org/10.48550/ARXIV.2212.02581

Crick, F. H. C., & Watson, J. D. (1954). The complementary structure of deoxyribonucleic acid. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, *223*(1152), 80–96. https://doi.org/10.1098/rspa.1954.0101

Crismore, A., & Vande Kopple, W. J. V. (1997a). Hedges and Readers: Effects on Attitudes and Learning. In *Hedging and Discourse* (pp. 83–114). de Gruyter. https://doi.org/10.1515/9783110807332.83

Crismore, A., & Vande Kopple, W. J. V. (1997b). The effects of hedges and gender on the attitudes of readers in the United States toward material in a science textbook. In *Culture and Styles of Academic Discourse* (pp. 223–247). de Gruyter. https://doi.org/10.1515/9783110821048.223

Demuth, R., Parchmann, I., & Ralle, B. (Eds.). (2006). *Chemie im Kontext: Sekundarstufe II* (1st ed.). Cornelsen.

Fang, Z. (2005). Scientific literacy: A systemic functional linguistics perspective. *Science Education*, *89*(2), 335–347. https://doi.org/10.1002/sce.20050

Goldman, S. R., & Bisanz, G. L. (2002). Toward a Functional Analysis of Scientific Genres: Implications for Understanding and Learning Processes. In J. Otero, J. A. León, & A. C. Graesser (Eds.), *The Psychology of Science Text Comprehension* (1st ed., pp. 19–50). Erlbaum.

Hagen, R., Watermann, R., & Nückles, M. (2022). Der Einfluss wissenschaftlicher Textgenres auf das Textverständnis und epistemologische Überzeugungen bei Studierenden der Bildungswissenschaften. *Swiss Journal of Educational Research*, *44*(2), 237–249. https://doi.org/10.24452/sjer.44.2.6

Hampp, R. (2022). Hedging: Zum kommunikativen Verständnis zwischen schreibenden Wissenschaftler*innen und rezipierenden Student*innen. *zeitschrift für interdisziplinäre schreibforschung*, 81–96. https://doi.org/10.48646/ZISCH.220605

Harmon, J. E., & Gross, A. G. (2004). *The Scientific Article: From Galileo's New Science to the Human Genome*. https://fathom.lib.uchicago.edu/2/21701730/

Hasty, P., Bradley, A., Morris, J. H., Edmondson, D. G., Venuti, J. M., Olson, E. N., & Klein, W. H. (1993). Muscle deficiency and neonatal death in mice with a targeted mutation in the *myogenin* gene. *Nature*, *364*(6437), Article 6437. https://doi.org/10.1038/364501a0

Hill, S. S., Soppelsa, B. F., & West, G. K. (1982). Teaching ESL Students to Read and Write Experimental-Research Papers. *TESOL Quarterly*, *16*(3), 333–347. https://doi.org/10.2307/3586633

Horn, K. (2001). The Consequences of Citing Hedged Statements in Scientific Research Articles. *BioScience*, *51*(12), 1086–1093. https://doi.org/10.1641/0006-3568(2001)051[1086:TCOCHS]2.0.CO;2

Hyland, K. (1994). Hedging in academic writing and EAF textbooks. *English for Specific Purposes*, *13*(3), 239–256. https://doi.org/10.1016/0889-4906(94)90004-3

Hyland, K. (1996). Talking to the Academy. Forms of Hedging in Science Research Articles. *Written Communication*, *13*(2), 251–281. https://doi.org/10.1177/0741088396013002004

Hyland, K. (2000). Hedges, Boosters and Lexical Invisibility: Noticing Modifiers in Academic Texts. *Language Awareness*, *9*(4), 179–197. https://doi.org/10.1080/09658410008667145

Irmer, E., Mihlan, M., & Töhl-Borsdorf, J. (2010). *Elemente Chemie 7-10 Niedersachsen G8* (1st ed.). Ernst Klett Verlag.

Jaek, A., Kinzel, C., Kronabel, C., & Peters, J. (2017). *Fokus Chemie SI Gesamtband* (1st ed.). Cornelsen.

Jensen, J. D. (2008). Scientific Uncertainty in News Coverage of Cancer Research: Effects of Hedging on Scientists' and Journalists' Credibility. *Human Communication Research*, *34*(3), 347–369. https://doi.org/10.1111/j.1468-2958.2008.00324.x

Kassambara, A. (2023). *rstatix: Pipe-friendly framework for basic statistical tests* [Manual]. https://cran.r-project.org/package=rstatix

Kraft, L., Ratermann, M., & Stoppel, F. (Eds.). (2015). *NEO Chemie SI Niedersachsen Gesamtband*. Schroedel.

Kuhlmann, J. W., Junius, M., Diederichsen, U., & Steinem, C. (2017). SNARE-Mediated Single-Vesicle Fusion Events with Supported and Freestanding Lipid Membranes. *Biophysical Journal*, *112*(11), 2348–2356. https://doi.org/10.1016/j.bpj.2017.04.032

Kund, J., Kruse, J., Gruber, A., Trentin, I., Langer, M., Read, C., Neusser, G., Blaimer, D., Rupp, U., Streb, C., Leopold, K., Schacher, F. H., & Kranz, C. (2023a). Multimodale Analyse lichtinduzierter Wasseroxidation in nanoporösen Blockcopolymer-Membranen**. *Angewandte Chemie*, *135*(22), e202217196. https://doi.org/10.1002/ange.202217196

Kund, J., Kruse, J.-H., Gruber, A., Trentin, I., Langer, M., Read, C., Neusser, G., Blaimer, D., Rupp, U., Streb, C., Leopold, K., Schacher, F. H., & Kranz, C. (2023b). Multimodal Analysis of Light-Driven Water Oxidation in Nanoporous Block Copolymer Membranes**. *Angewandte Chemie International Edition*, *62*(22), e202217196. https://doi.org/10.1002/anie.202217196

Lakoff, G. (1973). Hedges: A Study In Meaning Criteria And The Logic Of Fuzzy Concepts. *Journal of Philosophical Logic*, *2*(4), 458–508. https://doi.org/10.1007/BF00262952

Lewin, B. A. (2005). Hedging: An exploratory study of authors' and readers' identification of 'toning down' in scientific texts. *Journal of English for Academic Purposes*, *4*(2), 163–178. https://doi.org/10.1016/j.jeap.2004.08.001

Liang, G., Sadhukhan, T., Banerjee, S., Tang, D., Zhang, H., Cui, M., Montesdeoca, N., Karges, J., & Xiao, H. (2023a). Reduction of Platinum(IV) Prodrug Hemoglobin Nanoparticles with Deeply Penetrating Ultrasound Radiation for Tumor-Targeted Therapeutically Enhanced Anticancer Therapy. *Angewandte Chemie International Edition*, *62*(22), e202301074. https://doi.org/10.1002/anie.202301074

Liang, G., Sadhukhan, T., Banerjee, S., Tang, D., Zhang, H., Cui, M., Montesdeoca, N., Karges, J., & Xiao, H. (2023b). Reduktion von Platin(IV)-Prodrug Hämoglobin Nanopartikeln mit tief eindringender Ultraschallbestrahlung für eine gezielte und therapeutisch verstärkte Krebstherapie. *Angewandte Chemie*, *135*(22), e202301074. https://doi.org/10.1002/ange.202301074

Low, G. (1996). Intensifiers and Hedges in Questionnaire Items and the Lexical Invisibility Hypothesis. *Applied Linguistics*, *17*(1), 1–37. https://doi.org/10.1093/applin/17.1.1

Markl, J. (Ed.). (2018). *Markl Biologie Oberstufe*. Ernst Klett Verlag.

Mourez, M., Kane, R. S., Mogridge, J., Metallo, S., Deschatelets, P., Sellman, B. R., Whitesides, G. M., & Collier, R. J. (2001). Designing a polyvalent inhibitor of anthrax toxin. *Nature Biotechnology*, *19*(10), 958–961. https://doi.org/10.1038/nbt1001-958

Patalag, L. J., Sibold, J., Schütte, O. M., Steinem, C., & Werz, D. B. (2017). Gb_3 Glycosphingolipids with Fluorescent Oligoene Fatty Acids: Synthesis and Phase Behavior in Model Membranes. *ChemBioChem*, *18*(21), 2171–2178. https://doi.org/10.1002/cbic.201700414

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825–2830.

Phillips, L. M., & Norris, S. P. (2009). Bridging the Gap Between the Language of Science and the Language of School Science Through the Use of Adapted Primary Literature. *Research in Science Education*, *39*(3), 313–319. https://doi.org/10.1007/s11165-008-9111-z

Pointal, L. (2019). *Treetaggerwrapper 2.3* [Computer software]. https://treetaggerwrapper.readthedocs.io/

R Core Team. (2023). *R: A language and environment for statistical computing* [Manual]. R Foundation for Statistical Computing. https://www.r-project.org/

Schiller, A., Teufel, S., Stöckert, C., & Thielen, C. (1999). *Guidelines für das Tagging deutscher Textcorpora mit STTS. (Kleines und großes Tagset)*. https://www.ims.uni-stuttgart.de/documents/ressourcen/lexika/tagsets/stts-1999.pdf

Schmid, H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. *Proceedings of International Conference on New Methods in Language Processing*, *12*.

Schmid, H. (1995). Improvements in Part-of-Speech Tagging with an Application to German. *Proceedings of the ACL SIGDAT-Workshop*.

Schneck, F., Schendzielorz, F., Hatami, N., Finger, M., Würtele, C., & Schneider, S. (2018). Photochemically Driven Reverse Water-Gas Shift at Ambient Conditions mediated by a Nickel Pincer Complex. *Angewandte Chemie International Edition*, *130*(44), 14690–14695. https://doi.org/10.1002/anie.201803396

Schneider, A., Curado, C., Lystbaek, T. B., Osuna, S., & Hauer, B. (2023a). Harnessing the Structure and Dynamics of the Squalene-Hopene Cyclase for (−)-Ambroxide Production**. *Angewandte Chemie International Edition*, *62*(22), e202301607. https://doi.org/10.1002/anie.202301607

Schneider, A., Curado, C., Lystbaek, T. B., Osuna, S., & Hauer, B. (2023b). Verwendung der Struktur und Dynamik der Squalen-Hopen-Zyklase für die (−)-Ambroxid Produktion**. *Angewandte Chemie*, *135*(22), e202301607. https://doi.org/10.1002/ange.202301607

Schröder, H. (1998). „Ich sage das einmal ganz ungeschützt“ – Hedging und wissenschaftlicher Diskurs. In L. Danneberg & J. Niederhauser (Eds.), *Darstellungsformen der Wissenschaften im Kontrast: Aspekte der Methodik, Theorie und Empirie* (pp. 263–276). Gunter Narr.

Sifa, R., Pielka, M., Ramamurthy, R., Ladi, A., Hillebrand, L., & Bauckhage, C. (2019). Towards Contradiction Detection in German: A Translation-Driven Approach. *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2497–2505. https://doi.org/10.1109/SSCI44817.2019.9003090

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Routledge.

Takimoto, M. (2015). Assertions and lexical invisibility in EFL learners' academic essays. *Journal of Pragmatics*, *89*, 85–99. https://doi.org/10.1016/j.pragma.2015.09.009

Tarazona, J. V., Court-Marques, D., Tiramani, M., Reich, H., Pfeil, R., Istace, F., & Crivellente, F. (2017). Glyphosate toxicity and carcinogenicity: A review of the scientific basis of the European Union assessment and its differences with IARC. *Archives of Toxicology*, *91*(8), 2723–2743. https://doi.org/10.1007/s00204-017-1962-5

Vincze, V., Szarvas, G., Farkas, R., Móra, G., & Csirik, J. (2008). The BioScope corpus: Biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics*, *9*(11), Article 11. https://doi.org/10.1186/1471-2105-9-S11-S9

von Hoff, E. (2020). *Entwicklung und Evaluation von Konzepten und Formaten zum Wissenstransfer von der Forschung in die Schule und Öffentlichkeit—Am Beispiel des SFB 803* [Phd, Georg-August-Universität Göttingen]. https://doi.org/10.53846/goediss-8269

von Hoff, E., Aljets, H., Milsch, N., Waitz, T., & Mey, I. (2019). Arbeiten mit Adaptierter Primärliteratur im Chemieunterricht. *CHEMKON*, *26*(5), 190–195. https://doi.org/10.1002/ckon.201800070

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer. https://doi.org/10.1007/978-3-319-24277-4

Wonham, M. J., de-Camino-Beck, T., & Lewis, M. A. (2004). An epidemiological model for West Nile virus: Invasion analysis and control applications. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *271*(1538), 501–507. https://doi.org/10.1098/rspb.2003.2608

Wurl, F., Stipurin, S., Kollar, J. I., & Strassner, T. (2023a). Synthese und photophysikalische Eigenschaften monometallischer C^C* Platin(II)-Formamidinatkomplexe mit sterisch anspruchsvollen Liganden. *Angewandte Chemie*, *135*(22), e202301225. https://doi.org/10.1002/ange.202301225

Wurl, F., Stipurin, S., Kollar, J. I., & Strassner, T. (2023b). Synthesis and Photophysical Properties of Monometallic C^C* Platinum(II) Formamidinate Complexes using Sterically Demanding Ligands. *Angewandte Chemie International Edition*, *62*(22), e202301225. https://doi.org/10.1002/anie.202301225

Yarden, A. (2009). Reading Scientific Texts: Adapting Primary Literature for Promoting Scientific Literacy. *Research in Science Education*, *39*(3), 307–311. https://doi.org/10.1007/s11165-009-9124-2

Yarden, A., Brill, G., & Falk, H. (2001). Primary literature as a basis for a high-school biology curriculum. *Journal of Biological Education*, *35*(4), 190–195. https://doi.org/10.1080/00219266.2001.9655776

Yarden, A., Phillips, L. M., & Norris, S. P. (2015). *Adapted Primary Literature: The Use of Authentic Scientific Texts in Secondary Schools* (1st ed., Vol. 22). Springer. https://doi.org/10.1007/978-94-017-9759-7

Zhang, Y., Zhao, C., Ma, C., Cai, Z., Trienes, S., & Ackermann, L. (2023a). Photocatalytic C−C Cleavage of Methylenecyclobutanes for γ,δ-Unsaturated Aldehydes by Strain Release. *Angewandte Chemie International Edition*, *62*(22), e202300166. https://doi.org/10.1002/anie.202300166

Zhang, Y., Zhao, C., Ma, C., Cai, Z., Trienes, S., & Ackermann, L. (2023b). Photokatalytische Ringöffnung von Methylencyclobutanen zur Synthese von *γ,δ* -ungesättigten Aldehyden. *Angewandte Chemie*, *135*(22), e202300166. https://doi.org/10.1002/ange.202300166

Authenticity of scientific literature – Analyzing the usage of hedging in adapted scientific literature

Statements and Declarations

Funding

The authors declare that no funding was received for this work.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose

Authors' contributions

Hilko Aljets designed the study, collected data, performed analyses and wrote the manuscript. Thomas Waitz supervised and oversaw the entire study procedure. All authors revised and approved the final version of the manuscript. The results of this study have already been partially published in Hilko Aljets' dissertation.

Acknowledgements

We would like to thank Ingo Mey for fruitful discussions and advice on the study as well as all authors of the adapted articles.

Data availability

The ASL articles analyzed in German are available on request. The BioScope dataset used for training is publicly available from Vincze et al. (2008) or at https://rgai.inf.u-szeged.hu/node/105.

Code availability

The Python scripts and lists of features can be downloaded from https://gitlab.gwdg.de/hilko.aljets1/hedging.