



UNIVERSITY OF HELSINKI



<https://helda.helsinki.fi>

Helda

---

## When the unconventional becomes convention : epistemic stance in English as a lingua franca research articles

Shchemeleva, Irina

De Gruyter Mouton

2022-06-13

---

Shchemeleva, I 2022, 'When the unconventional becomes convention : epistemic stance in English as a lingua franca research articles', Journal of English as a Lingua Franca, vol. 11, no. 1, pp. 1-23. <https://doi.org/10.1515/jelf-2022-2075>

---

<http://hdl.handle.net/10138/358600>

10.1515/jelf-2022-2075

---

unspecified

publishedVersion

---

*Downloaded from Helda, University of Helsinki institutional repository.*

*This is an electronic reprint of the original article.*

*This reprint may differ from the original in pagination and typographic detail.*

*Please cite the original version.*

Irina Shchemeleva\*

# When the unconventional becomes convention: epistemic stance in English as a lingua franca research articles

<https://doi.org/10.1515/jelf-2022-2075>

**Abstract:** In the last decade numerous studies have been published in the field of English as an academic lingua franca (academic ELF). The majority of them, however, are devoted to the speaking mode. The present paper investigates written academic ELF texts – research articles in the social sciences and humanities. It explores the ways in which L2 speakers express epistemic stance in their texts and analyzes ELF-specific non-conventional forms of epistemic stance expressions. I use data from the SciELF corpus (<https://www2.helsinki.fi/en/researchgroups/english-as-a-lingua-franca-in-academic-settings/research/welfa-corpus/scielf-corpus>), which is unique in that it allows the study of academic texts written by L2 speakers in their unedited form. To uncover whether non-conventional expressions are also used in published academic texts, and to gain a deeper understanding of the extent to which variability in language use is accepted in published texts, I use the Google Scholar database as a reference corpus. My analysis shows that L2 authors creatively exploit linguistic resources to express epistemic stance, demonstrating considerable variability in language use. Findings also indicate that some non-conventional expressions are found in published texts.

**Keywords:** academic writing; approximation; English as an academic lingua franca; epistemic stance; non-conventional expressions

**Аннотация:** В последние десятилетия изучение английского языка как языка международного общения в научной среде привлекает многих исследователей. Однако большинство работ посвящено изучению устного общения. Данная статья рассматривает письменные тексты, написанные авторами, для которых английский не является родным. В работе рассматриваются способы выражения эпистемической позиции в текстах научных статей по социальным и гуманитарным наукам и анализируются нестандартные формы выражения эпистемической позиции, специфичные для английского

---

\*Corresponding author: Irina Shchemeleva, HSE University, St. Petersburg, Russia; and University of Helsinki, Helsinki, Finland, E-mail: ishemeliova@hse.ru

языка как языка международного общения. Работа выполнена на материале корпуса SciELF, уникальность которого заключается в том, что он позволяет изучать тексты, написанные не-носителями языка в исходном, неотредактированном виде. Для того, чтобы выявить, встречаются ли нестандартные формы в опубликованных текстах и сделать выводы о том, насколько допускается вариативность в языковом употреблении в письменных научных текстах, используется корпус Google Scholar. Результаты анализа показывают, что не-носителя английского языка используют вариативные формы для выражения эпистемической позиции в научных текстах. При этом некоторые нестандартные формы встречаются в опубликованных текстах.

**Ключевые слова:** английский язык как язык международного общения в академической среде; научный текст; эпистемическая позиция; нестандартные выражения

## 1 Introduction

Over the last several years there has been a great increase in the proportion of written academic texts produced internationally by L2 speakers<sup>1</sup> of English compared to texts written by L1 speakers (see Hyland 2016). This expansion poses new challenges for linguists to study the characteristics of academic texts from the perspective of academic ELF, and to see whether ELF has any influence on English academic writing.

The present research focuses on a particular characteristic of academic writing – epistemic stance – in research articles (RAs) written by L2 speakers. Epistemic stance has been chosen because it is one of the immanent features of the RA genre: to be successful in their argumentation, writers have to take a certain position in respect to knowledge they present. Epistemic stance expressions in academic texts written by non-native speakers have been widely studied within English for academic purposes (EAP) research (e.g., Hu and Cao 2011; Vassileva 2001; Vold 2006). In many cases, the aim of such studies is to compare and contrast epistemic stance expressions of non-native speakers with a certain L1 background with those used by native speakers. As a rule, such research has a firm orientation towards normative usage of English by native speakers. By contrast, ELF research is not concerned with comparing native and non-native usage of English, but rather with establishing commonalities in the use of English in international

---

<sup>1</sup> I use the term ‘speaker’ to refer to those who produce both spoken and written language.

academic communication irrespective of speakers' L1 backgrounds (Mauranen et al. 2016). This study, conducted from an ELF perspective, aims to identify ways in which L2 speakers express epistemic stance in academic texts.

In academic ELF, epistemic stance expressions have been studied in oral communication (Metsä-Ketelä 2006, 2012; Mortensen 2010), but they have not been analyzed in written ELF texts on a large database.<sup>2</sup> I use the data from the SciELF corpus (SciELF 2015), which uniquely allows us to study academic texts written by L2 speakers before they are checked by any language brokers. The SciELF corpus is part of the first publicly available corpus of written academic ELF – WrELFA (<http://www.helsinki.fi/englanti/elfa/wrelfa.html>). To draw conclusions on whether the non-conventional epistemic stance expressions that can be identified in the ELF data are also used in published texts, I use the Google Scholar (GS) database as a reference point, thus comparing non-published texts written by L2 speakers from the SciELF corpus to published texts written by any speaker of English from the GS database.

By looking into the ways L2 speakers, irrespective of their L1 backgrounds, express epistemic stance in their academic texts, the present paper aims to answer the following questions: What are the alternative ways of expressing epistemic stance characteristic of academic writing in ELF? Are these ELF-specific, non-conventional epistemic stance expressions used in published academic texts? The answers to these questions will not only allow us to identify what linguistic resources authors use and how creatively they exploit these resources when expressing epistemic stance, but also help us uncover some processes that are taking place in academic ELF writing in general.

## 2 Academic ELF research

The studies on ELF have shown that in ELF communication we observe not only the contact between similects – L1-based lects that have recognizable features and that are known, e.g., as ‘Spanglish’, ‘Dunglish’ or ‘Finglish’ (Mauranen 2012: 29), but also a complex interaction of individual multilingual repertoires of speakers (Cogo 2017). Multilingualism is considered an inherent characteristic of ELF: Jenkins (2015) has developed the notion of English as a “*multilingua franca*” stressing the importance of different languages that come in contact with each

---

<sup>2</sup> To my knowledge, the only study that is devoted to epistemic stance expressions in written academic ELF (ELFA) is that of Mur Dueñas (2016), which contrastively analyzes hedging modal verbs in RAs in the field of Business Management written by L2 speakers from a variety of L1s and native English speakers.

other in ELF interaction, thus foregrounding the multilingualism of ELF communicative settings.

One of the domains in which English has firmly established itself as a *lingua franca* is academia. Unsurprisingly, then, research on academic ELF, especially its speaking mode, has become a distinct strand of ELF research. Mauranen's studies of academic ELF have demonstrated that the main driver of lexico-grammatical change in ELF is the process of approximation, when speakers use rough equivalents of target elements whose meaning is easily recognizable (Mauranen 2012). This is especially evident in the use of multi-word units. By a multi-word unit Mauranen means a "recurrent unit of language which is associated with a recognizable meaning and includes more than one element; it includes grammatical as well as lexical components, of which some are fixed and some may be variable" (Mauranen 2012: 141). Mauranen demonstrates that very often in ELF communication such multi-word units have a preposition or article different from standard English phrases, sometimes lexical or structural substitutions occur (e.g., *from other point of view, does it make any different*). However, despite these semantic or structural irregularities, they are easily recognizable as units of meaning (Mauranen 2012, 2015). Certain approximate forms become so commonplace in ELF communication that it is possible to speak of the process of *fixing*, when an approximate form is repeatedly used for a given meaning across different L1 speakers (Vetchinnikova 2014). Describing new emerging norms in ELF, Jenkins states that ELF speakers "innovate in English making full use of their multilingual resources to create their own preferred forms" (Jenkins 2011: 928). Although the conclusions about creativity and variability of L2 speakers' academic ELF usage are mostly drawn from research on spoken interaction in academic ELF, we might infer that some of the innovative forms found in speaking will gradually penetrate into written language.

Despite the ever-growing number of written academic texts produced internationally by L2 speakers of English, research on the written mode of academic ELF is still scarce. This is probably due to the fact that writing, compared to speaking, is generally more conventional and less accepting of change. The existing research on texts written by L2 speakers mostly rely on text corpora representing speakers of the same similect (e.g., Ingvarsdóttir and Arnbjörnsdóttir 2013; Mur Dueñas 2013). Recently, however, some studies on written academic ELF texts in general have appeared (see, e.g., Carey 2013; Lorés-Sanz 2016; Mur Dueñas 2016; Rozycki and Johnson 2013). These studies suggest that some processes typical of ELF speech (simplification, approximation) can be observed in writing. Mauranen (2020) shows that approximation, especially in the use of prepositions and articles,

is found in writing. In her analysis of published working papers, Anderson (2010) identified irregularities in the positioning of adverbs and in tense choice, along with variability in the use of prepositions and articles. On the basis of these findings, Anderson expressed the idea of a gradual ‘remodeling’ of written language. The study of Rozycki and Johnson (2013), who analyzed award-winning papers written by L2 speakers in the field of engineering, suggests that some variability in language use is already accepted in published papers. The authors identified non-canonical usage of articles, verbs, prepositions, adverbs and adjectives; discord in subject-verb and determinant-noun number (Rozycki and Johnson 2013: 161–163).

Similar conclusions are drawn by Tribble (2017), whose work did not specifically aim to study non-conventional forms, but rather to demonstrate that lexico-grammatical variations at clause-level are acceptable only because “the texts themselves conform strongly to genre expectations at stage and move level” (Tribble 2017: 38). Nevertheless, his analysis of research paper excerpts from one scientific journal in biology, whose authors were identified as non-native speakers of English, confirms the hypothesis of significant variability in language use in written scientific communication. Tribble points to a high level of non-canonical forms, i.e., forms with “low probability occurrences which contravene the canonical rules of traditional text book grammar” (Tribble 2017: 37) such as missing determiner, collocating preposition problem, unclear wording, etc., approximately one non-canonical use for every 60 words in his 6,239 word corpus (Tribble 2017: 37–39).

To my knowledge, the only study that specifically aims to investigate the presence of ELF and non-native English forms in published RAs in international journals from a variety of disciplinary fields is that of Martinez (2018). It provides evidence that lexical features of ELF are found in published texts. Moreover, by comparing two periods of time (2000–2005 and 2010–2015) Martinez concludes that some lexical items characteristic to ELF (e.g., plural forms of nouns *work* and *research*; the phrase *it is interesting to mention/emphasize/estimate/gather* instead of *it is worth mentioning/emphasizing/estimating/gathering*; the phrase *in this context* as a synonym of either *in light of this* or *thus*, etc.) are becoming increasingly accepted (Martinez 2018: 40). However, as the author states, the research is only an exploratory study. Taking into consideration all the conclusions on variability and acceptance of non-conventional forms in published academic texts drawn from different studies, at this point we may hypothesize that with increasing numbers of academic texts written by L2 speakers of English, alternative non-conventional forms might become more accepted.

### 3 Epistemic stance

Over the last 40 years, there have been numerous studies on epistemic stance and the closely related notion of epistemic modality. Traditionally, epistemic modality has been defined from a position of formal logic that stresses the importance of the perceived truth value of the proposition. In his classic definition, Lyons states that an epistemically modal utterance is “an utterance in which the speaker explicitly qualifies his commitment to the truth of the proposition expressed by the sentence he utters” (Lyons 1977: 797). Similar definitions were given by Palmer (1986) and Coates (1987).

In later research on epistemic modality, the notion of *the truth of the proposition* was partially replaced by other concepts. Stubbs (1996), for example, uses *commitment to the proposition* as a broad term covering a continuum of modal and evidential meanings; Marín-Arrese argues that epistemic stance expresses “commitment to the validity of the information” (Marín-Arrese 2013: 414). Kärkkäinen in her study of epistemic stance shows that there is no “clear orientation of the participants to the truth of what they are saying”, rather, the speakers “assess something as more or less reliable, or express their belief that such and such is the case” (Kärkkäinen 2003: 18).

Regarding the spectrum of meanings covered by epistemic stance, recent research tends to treat epistemic stance as a category that includes two types of meanings: epistemic modal and evidential (Kärkkäinen 2003; Marín-Arrese 2013). Boye (2012) explicitly identifies a category of *epistemicity* as a supercategory covering two subcategories: *epistemic justification* (traditionally known as evidentiality) and *epistemic support* (traditionally known as epistemic modality). The inclusion of evidential meanings alongside epistemic modal meanings in the category of epistemic stance seems to be justified because evidential expressions may signal either reliability or non-reliability of the source of information, thus expressing different degrees of commitment to the proposition.

For the present paper the classification of epistemic stance meanings suggested by Biber et al. (1999) appears to be the most promising: they treat epistemic stance as a semantic category conveying various modal and evidential meanings that “are used to present speaker comments on the status of information in a proposition” (Biber et al. 1999: 972). Epistemic stance markers, according to Biber et al., express a writer’s “certainty (or doubt), actuality, precision or limitation”, or indicate “the source of knowledge or the perspective from which the information is given” (Biber et al. 1999: 972). This classification allows us to establish the range of linguistic means used to express different epistemic stance meanings in the data: I treated all the linguistic expressions that could be assigned

to one of the semantic categories identified by Biber et al. as epistemic stance markers. Moreover, the list of examples provided by the authors to illustrate each category can serve as a reference point for classifying epistemic stance markers into conventional or non-conventional ones.

## 4 Data description

The present study is based on the SciELF corpus (SciELF 2015), a 750,000-word database of scientific papers written by authors for whom English is an L2. It was compiled in 2015 at the University of Helsinki under the direction of Prof. Anna Mauranen. All papers included in the corpus are original scientific texts that have not been proofread or edited by any language professionals and/or native speakers of English. Most of the papers are final drafts of unpublished manuscripts.

The papers in the corpus follow a broad disciplinary division into Science (Sci) and Social Sciences and Humanities (SSH). Since the literature on academic discourse shows that there are significant variations across disciplines in the way authors express epistemic stance, especially between the so-called hard and soft sciences (Gray 2015; Hyland 1998, 2005; Varttala 2001), I concentrate on the latter and study the ways in which authors express epistemic stance in a broad academic domain of SSH. The SSH corpus comprises 70 RAs totalling 399,660 words. RAs included in the corpus cover a wide range of disciplines within social sciences (47%), humanities (35%), and behavioural sciences (18%).

One of the principles that guided the compilation of the SciELF corpus in general was consideration of L1 of the first author of the paper.<sup>3</sup> The aim was to have balanced samples of texts written by authors with different L1s, thus maximizing the variability of L1s, and to ensure that the corpus is representative in terms of what academic writing in ELF is like. The distribution of authors' L1 is given in Table 1; the overview of the academic roles of the first author's is presented in Table 2.

As can be seen from Table 2, the majority of authors are senior and junior researchers who aim to publish their papers in journals in English. Thus, the SSH corpus can be called a corpus of L2 use in international written scientific communication.

---

<sup>3</sup> We assume that the first author of the paper is responsible for the final text.

**Table 1:** Composition of the SSH corpus.

L1	No. of articles	No. of words	% of words	Avg. words/article
Chinese	10	44,196	11.06%	4,420
Czech	9	41,548	10.4%	4,662
Finnish	9	52,469	13.13%	5,830
French	8	48,373	12.11%	6,047
Italian	5	31,249	7.82%	6,250
Portuguese (Brazil)	6	39,223	9.81%	6,537
Romanian	4	25,197	6.3%	6,299
Russian	7	44,053	11.02%	6,293
Spanish	6	37,657	9.42%	6,276
Swedish	6	35,695	8.93%	5,949
Total	70	399,660	100%	5,856

**Table 2:** Authors' academic roles in the SSH corpus.

Role	No. of articles	No. of words	% of total words	Avg. words/article
Junior staff	43	232,501	58.17%	5,407
Senior staff	20	119,280	29.85%	5,964
Research student	6	42,009	10.51%	7,002
Unknown	1	5,870	1.47%	5,870

## 5 Methods

In order to determine the range of linguistic means used to express epistemic stance in the texts, the corpus was manually analyzed at the initial stage of the research. To ensure that the identified linguistic forms express epistemic stance according to established categories and to reduce subjectivity of interpretation, a second annotator, who is an experienced ELF researcher holding a PhD in linguistics, was invited to participate.

First, both annotators manually analyzed one paper from the corpus. In their analysis annotators relied on the lists of epistemic stance markers described in grammars of English (Biber et al. 1999; Huddleston et al. 2010 [2002]; Quirk et al. 1985) and in previous studies of epistemic stance, epistemic modality, evidentiality and similar phenomenon like hedges, boosters and evaluation (Aijmer and Simon-Vandenberg 2008; Hyland 1998; Varttala 2001; etc.). However, they did not restrict their analysis to these items: all linguistic expressions that conveyed epistemic stance meanings were marked.

Using the classification of epistemic stance markers by Biber et al. (1999), we marked every linguistic item that expressed an epistemic meaning as an epistemic stance marker:

- (1) *Some of these actions **might** involve a specific relationship ...* (SSH 46)
- (2) ***It is obvious** that some conditions are ...* (SSH12)
- (3) *Yet, he **seems** to neglect that the State also ...* (SSH18)

The results were compared and all instances were considered, including those marked by only one annotator. The differences mainly referred to the use of modal auxiliaries, especially *may* ([4]). We decided to include into the analysis all cases of the modal *may* where epistemic meaning can be implied:

- (4) *This paper will attempt to look into such linguistic means which, when in cooperation, **may create** a distinctive aesthetic effect.* (SSH16)

There were also cases of inconsistency in marking of some expressions (e.g., *mainly*). Sometimes evaluative adverbs (e.g., *most importantly*) were marked as epistemic by one annotator. Every such case was discussed and the agreement reached about inclusion in the sample.

Three additional papers were analyzed by both annotators to test the reliability of the annotating process. The kappa coefficient was 0.72, which is considered a substantial agreement (Landis and Koch 1977), and allows us to conclude that the criteria for identifying expressions are valid. After that, the whole corpus was manually analyzed and all epistemic stance expressions were marked.

In the analysis it was found that some stance markers are used as parts of larger wholes (multi-word units) whose meaning or form seem to be non-standard. For example, in (5) the stance expression – *has shown* – is absolutely conventional, while the reference to previous research – *the so far research* – is problematic (this collocation is not found either in grammars or collocation dictionaries; a conventional expression is *recent research*), and it renders the whole unit *the so far research has shown* slightly unidiomatic, or unconventional:

- (5) ***The so far research** (Firbas 1995 ...) **has shown** that the writer's communicative purpose may be ...* (SSH16)

In my analysis I take a holistic approach to cases similar to (5) and treat a multi-word unit as an epistemic stance expression since the epistemic meaning is expressed by the whole: in (5) it is the indication of the source of knowledge expressed by the whole phrase.

At the next stage of the analysis all identified markers were classified into conventional (i.e., those typical for Anglo-American norms of academic writing and included into English grammars and/or described in previous studies on epistemic stance) and non-conventional ones (i.e., those approximate items for which meaning is recognizable in the texts, e.g., *in the psychological viewpoint*). I used GS as a reference corpus to see whether the identified unconventional expressions are used in any published texts. It has been demonstrated that GS, which is not a linguistic corpus, but a large database of written academic texts, can be used in corpus studies of academic written texts (Brezina 2012; Flowerdew and Wang 2016; Heng Hartse and Kubota 2014). Brezina argues that it can be successfully searched “not only for academic *content* but [...] also for the *form* of academic expression” (Brezina 2012: 320, emphasis in the original). GS, in contrast to any native-speaker corpus, gives us an international perspective on the usage of English. Of course, we need to be aware that GS is not restricted to RAs, that it includes texts of different academic genres. Moreover, the quality of publications in the database is not checked, which means that the papers may come from very low impact, no impact and the so-called predatory publications. However, I believe that GS is still a valuable resource for the analysis since it shows the language use in different academic genres by different speakers across academia. The fact that it contains texts from all academic genres in all disciplines does not influence the results of my research since its aim is to see if an epistemic stance expression from the SSH corpus that I classified as non-conventional is used in published academic texts.

I searched GS for all epistemic stance markers that I had classified as non-conventional. If the search had fewer than 100 results, I considered this stance marker as an ‘individual non-conventional expression’; if there were more than 100 hits, I marked it as a ‘frequently found non-conventional expression’. Since my study is an essentially qualitative exploratory study, 100 was considered a good cut-off point for the reason of convenience: it is rather safe to say that if the expression occurs 100 times, it occurs frequently. Moreover, the GS search has shown that the expressions I searched for were either used in many fewer or many more than 100 papers.

Although GS is a useful tool for linguistic research, it does have some limitations. In some cases, which were very rare, it was not possible to draw conclusions on how often an expression is used in GS due to the high frequency of the phrase in various contexts and carrying different meanings compared to the meanings identified in the data. One example is the phrase *to him*, used in the corpus several times as an introductory phrase meaning *according to him*:

- (6) **To him**, it is expectable, and optimistic indeed, for the EU nations to further work on ... (SSH01)

GS returns more than 3 million hits, but they are used in other meanings.

Another example is the phrase *in \* perspective* used in GS more than 2.4 million times. An analysis of the first 500 hits in the search reveals that it is not used to convey the meaning ‘from somebody’s point of view’, as is found in sentence (7) from the data:

- (7) ***In our perspective***, any linguistic unit provides an instruction, whether lexical or grammatical. (SSH38)

A similar case is the phrase *to our view* (8) which is used in GS 24,100 times. The analysis of the first 500 usages, however, shows that this phrase is used as part of other expressions such as *according to our view*, *in opposition to our view*, *objection to our view*, etc.

- (8) ***To our view***, a historical themed festival is characterised by the provision of different services ... (SSH59)

In all such cases, I classified the expressions as individual non-conventional because I could not find evidence for how often the expressions are used in GS.

Finally, when all epistemic stance markers had been classified, I compiled a small sub-corpus in order to count the number of epistemic stance expressions in each category and to identify the proportions of each category of epistemic stance markers. The sub-corpus consists of 10 RAs written by speakers of 10 different L1 and belonging to 10 different disciplinary fields within SSH. Such a composition ensures the variability in terms of L1 and disciplinary representation, and makes it representative of the whole corpus. Although the proportion of epistemic stance markers in each category in the bigger corpus might be slightly different, I believe that the results of the counts of the expressions in the sub-corpus can provide us with a general idea of the distribution of epistemic stance markers in the whole corpus.

## 6 Results

The research demonstrates that L2 speakers use various means to express epistemic stance in their texts, both conventional and non-conventional.

### 6.1 Conventional epistemic stance markers

The analysis of the sub-corpus that was compiled to give an idea of the distribution of the expression in the whole SSH corpus demonstrates that in the overwhelming majority of cases L2 writers employ conventional epistemic stance expressions (Table 3). The high percentage of conventional epistemic stance expressions

**Table 3:** Distribution of epistemic stance markers in the sub-corpus.

	Conventional expressions	Non-conventional expressions		Total
		Individual	Frequently found	
No. of cases	604	18	2	624
Percentage	96.8%	2.9%	0.3%	100%

signals that in the current dataset L2 speakers are proficient users of English who construct their texts according to the norms of the RA genre. It also reflects a clear orientation to the native-speaker norms imposed by many journals, with publishers typically advising that manuscripts submitted for publication ought to be proofread, usually by a native speaker of English.

Conventional expressions found in the data are the same as described by Biber et al. (1999: 972–974). They are adverbials, modal verbs, cognition verbs, epistemic adjectives, stance nouns, etc. (see, e.g., [1]–[3]).

## 6.2 Non-conventional epistemic stance expressions

The calculations in the sub-corpus show that non-conventional expressions constitute a little more than 3% of all epistemic stance markers. Still, the very fact that they exist confirms the idea that L2 speakers are creative at making new non-conventional forms. In the SSH corpus I have found 132 non-conventional epistemic stance expressions that are used in 45 papers (64.3%).

Table 4 demonstrates that non-conventional epistemic stance expressions are used by multiple authors with different L1s. There does not seem to be any correlation between the L1 of the author and the number of non-conventional forms: 34 papers that present all 10 L1s from the corpus contain from one to three non-conventional expressions.

Of course, some authors do have individual specific features. For example, in one text we find three non-conventional variants of the expression *to some extent* (SSH63). However, these cases are rare. The most important finding of the analysis is that authors with various L1s approximate conventional epistemic stance expressions, making new forms.

Using the terminology of Mauranen (2012), we can say all the identified expressions fall into two broad categories: semantic and structural approximation, with several border cases combining irregularities in both meaning and form (Table 5).

**Table 4:** Distribution of non-conventional epistemic stance expressions in SSH corpus.

No. of cases	No. of articles	L1 of the author
10	1	Spanish
7	3	French, Portuguese, Spanish
6	2	Chinese, Czech
5	1	Finnish
4	5	French, Russian, Spanish, Swedish (2)
3	9	Chinese, Czech, French (3), Italian, Russian (2), Swedish
2	13	Chinese (2), Czech (3), Finnish (2), French, Portuguese, Romanian, Russian (2), Swedish
1	12	Czech (3), Finnish (3), Italian, Portuguese (2), Romanian (2), Russian
0	25	Chinese (6), Czech, Finnish (4), French, Italian (3), Portuguese (2), Romanian, Russian (2), Spanish (3), Swedish (2)

**Table 5:** Type of approximation.

	Semantic	Structural		Structural + semantic	Total
		Morphological	Syntactic		
No. of cases (%)	48 (36.4%)	11 (8.3%)	69 (52.3%)	4 (0.3%)	132 (100%)
No. of articles	29	8	34	4	45

In the next sections I shall first characterize non-conventional epistemic stance expressions in terms of the mechanisms underlying their formation, i.e., semantic and structural approximation, before moving onto a consideration of those non-conventional expressions that were found in both my corpus and the GS corpus.

### 6.2.1 Semantic approximation

In semantic approximation the speaker substitutes one or several elements of a multi-word unit on the basis of similarity in meaning, making the expression non-conventional but still easily recognizable. Example (9) illustrates semantic approximation of the expression *research of other scholars*:

- (9) ***Research by other individuals has demonstrated***, however, that family management and strategy can be effective and successful ... (SSH14)

The reader understands the meaning of the expression even though the word choice may be considered not typical.

Another example of non-standard (from the semantic point of view) expression is the phrase *simple research shows*, used by one of the authors in the corpus:

- (10) ***Simple research shows*** that “a typical” minibus for the citizens of St. Petersburg is ... (SSH54)

This usage is rather interesting because it allows us to infer how the writer coins the new non-conventional phrase. In the same paper we find a similar expression *simple mathematics shows*:

- (11) However, ***simple mathematics shows*** that the income of a family whose members regularly use “marchroutkas” needs to be high enough. (SSH54)

(11) can be considered normative, although the phrase *simple mathematics shows* is not as frequent in published academic texts as *simple arithmetic shows* (232 vs. 1,150 cases in GS). We might imply that the writer first extrapolated the meaning from the idiom *simple arithmetic shows*, and then from that created *simple research shows*, which is not conventional in English. These examples may give us an idea of how the process of approximation works.

What is also worth mentioning is that similar non-conventional expressions are found in texts written by authors with different L1s. For example, (12) and (13) are approximations of the multi-word unit *recent research shows*, but in (12) the author’s L1 is Czech, while in (13), Finnish:

- (12) ***The so far research*** ... has shown (SSH16)

- (13) ***The former research*** ... showed (SSH29)

These are good examples of the variability employed by L2 speakers: they both replace one element *recent* by other similar in meaning but not typical for the expression elements *so far* and *former*. These cases demonstrate the process of complexification of the language use that is observed both in ELF speaking and writing (Mauranen 2012).

Another set of examples that merits attention are cases of reference to common knowledge, which are frequently found in academic texts. In the data we see variability in the use of these expressions:

- (14) ***It is of common acceptance*** nevertheless, at least so far, the idea of unity. (SSH62)

- (15) ... ***it is apparently agreed*** that awareness of the behaviour and intention to achieve a given goal are ... (SSH48)

- (16) ... were not so neoliberal **as common argumentation assumes**, as the state still had a strong role in Brazil ... (SSH25)

In the data I have also found cases where the author modifies conventional expressions, thereby adding some extra epistemic meaning. In (17) the contribution of the paper to the field is diminished by the expression *to shed a small light* which is used instead of the more traditional *to shed some light*:

- (17) ... this paper **sheds a small light** on the pervasive nature of this phenomenon which might hopefully contribute to the understanding of how memes work in general. (SSH18)

The author further intensifies his/her detachment from the proposition by the modal *might* and the adverb *hopefully*. This “harmonic combination” (Coates 1983) opens a dialogue space, allowing for alternative interpretations of the presented research results.

The next example (18) where the adverbial *admittedly* is used to intensify the phrase *plausible to argue*, also indicates the author’s detachment from the expressed claim, thus opening up a possibility for alternative interpretations:

- (18) **It is admittedly plausible to argue** that the Court’s perception of Islam is oversimplified and far from neutral ... (SSH26)

### 6.2.2 Structural approximation

Although some semantic irregularities have been found in the data, the cases of structural approximation are more plentiful. Structural approximation is found on two levels: morphological and syntactic.

On the level of morphology all the identified cases refer to the use of a singular noun in the plural form (*evidences, researches, works*):

- (19) **All those evidences together seem to suggest** a more dialogical scientific text towards consensus. (SSH44)
- (20) Similarly, **researches** carried out in Sweden and Norway show that ... (SSH62)

These examples show that in L2 writing we observe the same process of simplification that is described in speaking ELF research, i.e., the regularization of some forms.

On the level of syntax, we can find diverse cases of approximation. To demonstrate this point, I will start with phrases where the grammar structures are non-standard:

- (21) *Rapid changes in information technology .... create various opportunities for the born global phenomenon to grow, **as it is documented in previous research.*** (SSH15)
- (22) ***From the research of <NAME> there was found** that only 13% of selected sample ...* (SSH12)

Probably the most obvious cases of structural approximation are found in the use of prepositions ([23]–[26]):

- (23) ***To** Confucius, learning is an interconnected rather than intermittent process.* (SSH02)
- (24) ***To** Lakoff, an expression like over cannot be analysed ...* (SSH38)
- (25) *New European investors are often attracted to invest in Georgia **by** the following survey results of the World Bank* (SSH17)
- (26) ***In a psychological viewpoint,** emotions are regarded as intensive reactions of the body ...* (SSH37)

The examples above express the source of knowledge or the perspective from which the information is given; in (23)–(25) a more conventional preposition would be *according to*, in (26), *from*. All these cases suggest that approximate use of preposition is typical not only for academic ELF speech but also for written academic ELF, which might imply that we observe similar processes in language use in both spoken and written modes of academic ELF.

Another illustration of the similarity of processes involved in ELF speaking and writing are the cases of structural approximation in which the same or a similar non-conventional expression is used by authors with different L1s. For example, the following variants of the expression *from* \* *perspective* are found in the corpus:

- in this perspective* (French, Spanish, Russian);  
*in this jurisdictional perspective* (Italian);  
*since this mere perspective* (Spanish);  
*in our own perspectives* (French).

The last two examples combine structural and semantic approximation since the irregularities are observed both in grammar (prepositions) and in the word choice (*this mere*; *our own*). Another example showing that a non-conventional form can combine both semantic and structural approximation is (27).

- (27) ***The popular opinion has been that*** voluntary disclosure is closer to “green-washing” than to real transparency, and the present contribution partially supports this view. (SSH50)

Here the general meaning of the expression seems quite unusual: we expect that the author would contradict the proposition that is introduced by the phrase *the popular opinion*, which would normally refer to some ironic belief, but the context shows that the meaning is the opposite: the author’s research supports the idea.

Despite irregularities in meaning or in form, in all the examples given above the general meaning of the expression is easily understandable. I have found only one case, which might be considered an exception, whose meaning is not easy to identify. In (28) the exact meaning of the expression *we are thus convinced to hope* could not be stated: we might infer that it is either a transformation of *we hope to convince* or that it might mean that something convinces us that we should hope:

- (28) ***We are thus convinced to hope*** that in the long-term Karran’s *Magna Charta* would “acquire the same status and impact as the AAUP statement”. (SSH01)

I have classified this case as a ‘border case’ combining semantic and structural approximation.

All the examples in Sections 6.2.1 and 6.2.2 are individual non-conventional expressions that are found in GS less than 100 times. The next section presents the analysis of those expressions that are rather frequent in GS.

### 6.2.3 Frequently found non-conventional expressions

Probably the most interesting findings in the analysis of epistemic stance expressions were those expressions that seem unconventional from the point of view of standard grammar, but which turned out to be quite commonly used in academic texts indexed in GS. The two examples below contain the phrase *according to*. Both of them seem unconventional but are very often used: there are 14,150 cases in (29) and 1,960 cases in (30).

- (29) ***According to my view***, the shift signified the moral cultivation of actor’s character. (SSH27)
- (30) ***According to himself***, he is supposed to get married at the age of about 23, and feels that he has to be able to arrange a wedding ... (SSH68)

The data also contained several variants of the conventional expression *to some extent*, a phrase that is commonly found in academic texts (about 3 million hits in

GS). It normally expresses some imprecision. Probably due to its frequent usage, L2 writers use it with variability:

- (31) *In Russia LGBT community failed to form the common position and was **at large extent** divided by the state pressure.* (SSH58)
- (32) *The decisions of being member of specific organisations will depend **on a great extend** to the same observed and unobserved determinants of offering volunteer work.* (SSH63)
- (33) *The kidnapping plot is **to its greatest extent** reported and the discussion ...* (SSH70)

In (31) and (32) we can see approximation of prepositions, in (32) the word *extent* is misspelled; in (33) the word *greatest* is used instead of a more regular word *fullest* and the possessive pronoun *its* instead of a more conventional article *the*. Although they seem non-conventional, all three expressions are found in published texts: GS returns 670, 480 and 2,360 hits, respectively.

This category of frequently found non-conventional expressions is not large: in the whole corpus I have found 22 such cases. However, the very fact that such expressions exist gives another proof to the idea expressed by previous ELF research that some ELF-specific forms are becoming accepted in published texts.

## 7 Discussion

The data used in the research – the SciELF corpus– is exceptional because it provides an opportunity to study the authentic use of written academic ELF. It accords perfectly with the aims of the study and allows us to analyze how authors exploit linguistic recourses when they express epistemic stance, and to identify variability in language use characteristic to ELF. One of the most important features of the present study is that L2 texts are not compared to L1 texts: the GS database, which is not an L1 database, but a collection of published academic texts written by any speakers of English, is used to check whether the identified epistemic stance expressions are also found in other academic texts. By analyzing the language of the GS database we may get the international perspective on the use of English in academia and draw some conclusions on the current state of academic writing in English in general and on the acceptability of non-conventional language forms used in written ELF texts in particular. The choice of GS as a reference database has some limitations, of course, because it is not

specifically compiled for linguistic studies. However, since it is a huge collection of published academic texts of different genres, the GS search may be considered a valid tool for identifying if a certain unconventional expression is found in published texts. The cases when it was not possible to reach a clear conclusion about the frequency of the expression's usage were very rare.

The results of the analysis of epistemic stance expressions used in the data clearly show that in RAs written by L2 speakers there is considerable variability in the use of linguistic resources to express epistemic stance. The main challenge of the present research was to classify all identified expressions into conventional and non-conventional ones in order to study variability in the use of linguistic resources in written ELF texts. Reference to grammars and earlier studies on epistemic stance helped identify non-conventional expressions.

The study demonstrates that the vast majority of epistemic stance expressions are conventional (about 97% in the sub-corpus specially compiled for the frequency calculations). This is unsurprising, considering the fact that all authors wrote their papers in English with the aim of having them published, meaning they would have to follow a Standard English model of academic writing that generally equates to native speaker models. This is related to issues of normativity in the academic discourse imposed by gatekeepers and, as a consequence, to the position of L2 scholars in the English-dominant academia (see, e.g., the discussion of the “linguistic injustice” doctrine in Flowerdew [2008, 2019] and the “academic literacy” framework by Lillis and Curry [2010]). The discussion of L2 scholars' position in the contemporary academia, which is very often an ELF academia, is outside the scope of this research since the data I use – the SciELF corpus – allows studying the texts in their original form, without considering ‘text histories’ and without discussing any changes that the writers have to do in the process of publication. However, the results of the study signal that L2 speakers in their writing have a clear orientation on the normative (i.e., native-speaker) use of epistemic stance expressions. It suggests that L2 speakers have to invest a lot of efforts and resources into developing their ability to conform to normative academic writing in English.

The results also show that along with conventional expressions, L2 speakers use non-conventional epistemic stance markers. This confirms the hypothesis that alternative ways of expressing epistemic stance that are characteristic to ELF would be found in written texts: L2 speakers in their written academic texts approximate both in form and meaning, creating non-conventional expressions that are in the overwhelming majority of cases perfectly comprehensible. Of course, some expressions identified in the dataset can be described as individual authorial features, but the fact that these approximate forms are found in 64.3% papers from the corpus means that approximation, and hence, variability in the

language use, is not an individual trait of a certain L2 speaker, but a characteristic feature of academic ELF writing in general.

The analysis also shows that sometimes the same non-conventional expression is found in texts written by authors from different L1 backgrounds, which suggests that L2 speakers, irrespective to their L1, employ the same mechanisms of approximation to create non-conventional expressions. Moreover, these mechanisms seem to be the same as described in the studies of the speaking mode of ELF. Thus, the results support the ideas of ELF researchers who have suggested that the surface processes that are happening in academic ELF are similar in speaking and writing (Lorés-Sanz 2016; Mauranen 2016, 2020).

It was predictable from the very beginning that there would be some individual variability in the use of epistemic stance markers in written academic ELF texts since this variability is typical of ELF speech. However, we could not predict that some non-conventional forms used by L2 speakers in the data would also be found in published texts. The fact that the identified non-conventional approximate expressions are used in the GS database means that variability characteristic to ELF is becoming accepted in published texts. This conclusion coincides with the findings of Rozycki and Johnson (2013), Tribble (2017), and Martinez (2018), who analyzed published research articles written by L2 speakers of English. Moreover, the frequent use of such expressions in published texts included in GS indicates that some variability is already accepted and implies that some of these expressions may be in the process of fixing. However, to explore this further, diachronic studies should be conducted. At this point in time, we can only witness that some of these non-conventional forms are used more often than others and predict that highly conventional norms of English academic writing may be undergoing some changes under the influence of ELF writing. In general, the research supports the idea of a gradual ‘remodeling’ of the written language proposed by Anderson (2010); we can predict that the extent of the accepted variability would be growing as more and more L2 speakers write their academic texts in English, and thus individual expressions that are now rare may become more commonplace.

These conclusions are drawn on the basis of analysis of epistemic stance expressions only and need to be further corroborated by the analysis of other features of written academic texts (as suggested by Martinez 2018). The scope of further research should be widened to include other phenomena that are typical of academic writing. Together with the findings from the present exploratory study, this may highlight the processes that are going on in ELF writing and reveal if the conventions of academic writing are influenced by these processes. The question that further research should answer is to what extent, if any, are the norms of academic writing in English changing under the influence of academic texts produced by L2 speakers?

## References

- Aijmer, Karin & Anne-Marie Simon-Vandenbergen. 2008. *The semantic field of modal certainty: A corpus-based study of English adverbs*. Berlin & New York: Mouton de Gruyter.
- Anderson, Laurie. 2010. Standards of acceptability in English as an academic lingua franca: Evidence from a corpus of peer-reviewed working papers by international scholars. In Roberto Cargliero & Jennifer Jenkins (eds.), *Discourses, communities, and global Englishes*, 115–143. Peter Lang: Bern.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad & Edward Finegan. 1999. *The Longman grammar of spoken and written English*. London: Longman.
- Boye, Kasper. 2012. *Epistemic meaning: A crosslinguistic and functional-cognitive study* (Empirical Approaches to Language Typology 43). Berlin & Boston: Mouton de Gruyter.
- Brezina, Vaclav. 2012. Use of Google Scholar in corpus-driven EAP research. *Journal of English for Academic Purposes* 11(4). 319–331.
- Carey, Ray. 2013. On the other side: Formulaic organizing chunks in spoken and written academic ELF. *Journal of English as a Lingua Franca* 2(2). 207–228.
- Coates, Jennifer. 1983. *The semantics of the modal auxiliaries*. London: Croom Helm.
- Coates, Jennifer. 1987. Epistemic modality and spoken discourse. *Transactions of the Philological Society* 85(1). 110–131.
- Cogo, Alessia. 2017. ELF and multilingualism. In Jennifer Jenkins, Will Baker & Martin Dewey (eds.), *The Routledge handbook of English as a lingua franca*, 357–368. Abingdon, UK: Routledge.
- Flowerdew, John. 2008. Scholarly writers who use English as an additional language: What can Goffman's "stigma" tell us? *Journal of English for Academic Purposes* 7. 77–86.
- Flowerdew, John. 2019. The linguistic disadvantage of scholars who write in English as an additional language: Myth or reality. *Language Teaching* 52. 249–260.
- Flowerdew, John & Simon Ho Wang. 2016. Author's editor revisions to manuscripts published in international journals. *Journal of Second Language Writing* 32. 39–52.
- Gray, Bethany. 2015. *Linguistic variation in research articles: When discipline tells only part of the story*. Amsterdam & Philadelphia: John Benjamins.
- Heng Hartse, Joel & Ryuko Kubota. 2014. Pluralizing English? Variation in high-stakes academic texts and challenges of copyediting. *Journal of Second Language Writing* 24. 71–82.
- Hu, Guangwei & Feng Cao. 2011. Hedging and boosting in abstracts of applied linguistic articles: A comparative study of English- and Chinese-medium journals. *Journal of Pragmatics* 43. 2795–2809.
- Huddleston, Rodney Desmond, Geoffrey Keith Pullum & Laurie Bauer. 2010 [2002]. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Hyland, Ken. 1998. *Hedging in scientific research articles*. Amsterdam: John Benjamins Publishing.
- Hyland, Ken. 2005. Stance and engagement: A model of interaction in academic discourse. *Discourse Studies* 7(2). 173–192.
- Hyland, Ken. 2016. Academic publishing and the myth of linguistic injustice. *Journal of Second Language Writing* 31. 58–69.
- Ingvarsdóttir, Hafdís & Birna Arnbjörnsdóttir. 2013. ELF and academic writing: A perspective from the expanding circle. *Journal of English as a Lingua Franca* 2(1). 123–145.
- Jenkins, Jennifer. 2011. Accommodating (to) ELF in the international university. *Journal of Pragmatics* 43(4). 926–936.

- Jenkins, Jennifer. 2015. Repositioning English and multilingualism in English as a lingua franca. *Englishes in Practice* 2(3). 49–85.
- Kärkkäinen, Elise. 2003. *Epistemic stance in English conversation: A description of its interactional functions, with a focus on I think*. Amsterdam: John Benjamins.
- Landis, Jay Richard & Gary Grove Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33. 159–174.
- Lillis, Theresa & Mary Jane Curry. 2010. *Academic writing in a global context. The politics and practices of publishing in English*. London: Routledge.
- Lorés-Sanz, Rosa. 2016. ELF in the making? Simplification and hybridity in abstract writing. *Journal of English as a Lingua Franca* 5(1). 53–81.
- Lyons, John. 1977. *Semantics*, vol. 2. Cambridge: Cambridge University Press.
- Marín-Arrese, Juanna I. 2013. Stancetaking and inter/subjectivity in the Iraque inquiry: Blair versus Brown. In Juanna I. Marín-Arrese, Marta Carretero, Jorge Arús Hita & Johan van der Auwera (eds.), *English modality: Core, periphery and evidentiality* (Topics in English Linguistics [TiEL] 81), 412–445. Berlin & Boston: Mouton de Gruyter.
- Martínez, Ron. 2018. “Specially in the last years ...”: Evidence of ELF and non-native English forms in international journals. *Journal of English for Academic Purposes* 33. 40–52.
- Mauranen, Anna. 2012. *Exploring ELF: Academic English shaped by non-native speakers*. Cambridge: Cambridge University Press.
- Mauranen, Anna. 2015. What is going on in academic ELF? Findings and implications. In Paola Vettorel (ed.), *New frontiers in teaching and learning English*, 31–52. Cambridge: Cambridge Scholars Publishing.
- Mauranen, Anna. 2016. ELF and change. Paper presented at the International PhD conference on global Englishes, Southampton, UK, 24 June. <http://blog.soton.ac.uk/ilc/files/2016/07/AM240616.pdf> (accessed 4 May 2022).
- Mauranen, Anna. 2020. Good texts in non-standard English: ELF and academic writing. In Kumiko Murata (ed.), *ELF research methods and approaches to data and analyses: Theoretical and methodological underpinnings*, 57–80. Abingdon: Routledge.
- Mauranen, Anna, Niina Hynninen & Elina Ranta. 2016. English as an academic lingua franca. In Ken Hyland & Philip Shaw (eds.), *The Routledge handbook of English for academic purposes*, 65–77. Abingdon, Oxon: Routledge.
- Metsä-Ketelä, Maria. 2006. ‘Words are more or less superfluous’: The case of *more or less* in academic lingua franca English. *Nordic Journal of English Studies* 5(2). 117–143.
- Metsä-Ketelä, Maria. 2012. Frequencies of vague expressions in English as an academic lingua franca. *Journal of English as a Lingua Franca* 1/2. 263–285.
- Mortensen, Janus. 2010. *Epistemic stance marking in the use of English as a lingua franca*. Roskilde, Denmark: Roskilde Universitet electronic PhD Dissertation.
- Mur Dueñas, Pilar. 2013. Spanish scholars’ research article publishing process in English-medium journals: English used as a lingua franca? *Journal of English as a Lingua Franca* 2(2). 315–340.
- Mur-Dueñas, Pilar. 2016. Modal hedging verbs in English as a lingua franca (ELF) business management research articles. *Kalbotyra* 69. 153–178.
- Palmer, Frank R. 1986. *Mood and modality*. Cambridge: Cambridge University Press.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. London: Longman.
- Rozycki, William & Neil H. Johnson. 2013. Non-canonical grammar in Best Paper award winners in engineering. *English for Specific Purposes* 32(3). 157–169.

- SciELF. 2015. The SciELF corpus. Director: Anna Mauranen. Compilation manager: Ray Carey. <https://www2.helsinki.fi/en/researchgroups/english-as-a-lingua-franca-in-academic-settings/research/wrelfa-corpus/scielf-corpus> (accessed 7 May 2022).
- Stubbs, Michael. 1996. *Text and corpus analysis*. Oxford: Buckwell Publishers.
- Tribble, Christopher. 2017. ELFA versus Genre: A new paradigm war in EAP writing instruction? *Journal of English for Academic Purposes* 25. 30–44.
- Varttala, Teppo. 2001. *Hedging in scientifically oriented discourse: Exploring variation according to discipline and intended audience*. Tampere, Finland: University of Tampere Electronic PhD Dissertation.
- Vassileva, Irene. 2001. Commitment and detachment in English and Bulgarian academic writing. *English for Specific Purposes* 20(1). 83–102.
- Vetchinnikova, Svetlana. 2014. *Second language lexis and the idiom principle*. Helsinki: University of Helsinki Unpublished PhD Dissertation.
- Vold, Eva Thue. 2006. Epistemic modality markers in research articles: A cross-linguistic and cross-disciplinary study. *International Journal of Applied Linguistics* 16(1). 61–87.

## Bionote

### **Irina Shchemeleva**

HSE University, St. Petersburg, Russia

[ishemeliova@hse.ru](mailto:ishemeliova@hse.ru)

Irina Shchemeleva is Associate Professor and Dean of the School of Arts and Humanities at HSE University, Russia. Her current research interests are in the field of academic ELF and ERPP. She has published on topics related to L2 scholars' research writing practices and teaching academic skills at the tertiary level.