English as a Lingua Franca in research articles: the SciELF corpus

L’anglais lingua franca dans les articles de recherche : le corpus SciELF

Elizabeth Rowley-Jolivet

Electronic version
URL: http://journals.openedition.org/asp/4987
DOI: 10.4000/asp.4987
ISSN: 2108-6354

Publisher
Groupe d'étude et de recherche en anglais de spécialité

Printed version
Date of publication: 1 March 2017
Number of pages: 145-158
ISSN: 1246-8185

Electronic reference

This text was automatically generated on 1 May 2019.

Tous droits réservés
English as a Lingua Franca in research articles: the SciELF corpus

L’anglais lingua franca dans les articles de recherche : le corpus SciELF

Elizabeth Rowley-Jolivet

1. Introduction: English as the Lingua Franca of academia

Since the Second World War, English has become the lingua franca (ELF) of many domains, ranging from business to international politics to academia, and the globalization fostered by the rise of internet and new media has accentuated this development. The number of non-native speakers of English now greatly exceeds the number of native speakers. It is estimated for example, that fewer than one out of every four users of English in the world is a native speaker of the language (Crystal 2003), and as a result most interactions in English now take place in intercultural contexts, among non-native speakers of English or between native and non-native speakers. For ELF users, English is a means to communicate with speakers of another language worldwide, and not, or seldom, a means to achieve cultural and linguistic integration in an English L1 society. In this respect, the definition of ELF adhered to by the majority of researchers in the field distinguishes between ELF and English as a Foreign Language (EFL). While mainstream EFL generally posits that the goal of non-native speakers is to approximate the native variety, seen as the benchmark, as closely as possible, and considers that departures from this norm are deficiencies or errors, the result of L1 interference or poor competence (Belz 2002), ELF sees non-native Englishes as different rather than deficient, with their own emerging strategies and specific features. This approach has understandably generated considerable resistance from those who regard native speakers as the natural guardians of acceptable usage, as it implies that current changes in the language may be shaped at least as much by ELF users as by native speakers, and that the language may no longer belong to the latter, who now form a minority group of speakers.
As argued by Seidlhofer, among others, research into the characteristics of ELF is therefore necessary:

> In order for the concept of ELF to gain acceptance alongside English as native language, there have been calls for the systematic study of the nature of ELF—what it looks and sounds like and how people actually use it and make it work. (Seidlhofer 2005: 339–340)

One important domain that has adopted English as its common language, both in speech and in writing, is academia, as pointed out by Mauranen (2010: 7):

> Academia is in many ways a typical ELF domain: it is international, mobile and its dependence on English has skyrocketed in the last few decades. Academia is thoroughly dependent on cooperation across national borders and internationally negotiated standards, especially in science.

In order to understand present-day academic English, therefore, the object of study should arguably be lingua franca English, as it is this that characterizes the mainstream of academic English use, rather than native academic English (Mauranen et al. 2010). Moreover, in view of the prestige enjoyed by academic English in all knowledge fields, any changes in the language induced by the predominance of ELF users in academia may well become standard or accepted usage.

To study academic ELF, sufficiently large and well-constructed corpora are necessary. The present research note introduces a corpus – SciELF – which aims to enable this. SciELF is part of the WrELFA corpus (Written English as a Lingua Franca in Academic Settings), the written component of the larger ELFA1 project that covers both spoken and written academic ELF. The spoken component of ELFA has been available for research purposes since 2008, and has given rise to a large number of analyses and publications, while WrELFA became available in 2016. The project is directed by Anna Mauranen at the University of Helsinki.

To put WrELFA into context, I will first give a brief overview of ELF corpora, before describing in more detail in section 3 the different components of the WrELFA corpus, focusing in particular on SciELF. An illustration of the research and teaching interest of SciELF is given in the final section which looks at how a category of verbs that is very frequently needed in academic English, namely enabling verbs (allow, enable, permit), is used in academic ELF.

2. Brief overview of ELF corpora

<table>
<thead>
<tr>
<th>Full name</th>
<th>Date of completion</th>
<th>Size (in words)</th>
<th>Mode</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>VOICE</td>
<td>Vienna-Oxford International Corpus of English</td>
<td>2009</td>
<td>1 million w.</td>
<td>spoken</td>
</tr>
</tbody>
</table>

Table 1. The major spoken and written ELF corpora
A review of ELF research up to 2011 can be found in Jenkins et al. (2011); here I will just sketch in some of the major landmarks and available corpora. The date of 2000 can be taken as a convenient starting point for systematic studies of various features of ELF, as it saw the publication of Jenkins’ study of ELF pronunciation (Jenkins 2000). Jenkins identified what she called the Lingua Franca Core, i.e. the crucial pronunciation features that most often lead to miscommunication in EFL interactions, and the accommodation engaged in by ELF speakers to ensure intelligibility. She concluded that some features prominent in EFL teaching (such as the pronunciation of <th>) were not in fact crucial and that native English pronunciation was not the optimal model in ELF contexts. This was rapidly followed in 2001 by the launch of the first ELF spoken corpus on the initiative of Barbara Seidlhofer, the Vienna-Oxford International Corpus of English (VOICE; see Table 1). Completed in 2009, VOICE comprises one million words of transcribed spoken ELF from professional, educational and leisure domains, representing 120 hours of naturally occurring face-to-face interactions.

Two years later, in 2003, a second ELF spoken corpus was launched under the leadership of Anna Mauranen that focused on academic English, the corpus of English as a Lingua Franca in Academic Settings, or ELFA. The ELFA corpus was completed in 2008 and contains one million words of transcribed spoken academic ELF (approximately 131 hours of recorded speech). The recordings were made in various university settings in Finland. The latest addition to these spoken corpora is ACE, the Asian Corpus of English run by Andy Kirkpatrick from Hong Kong. Started in 2009, ACE was completed in 2014, and also contains one million words of naturally occurring, spoken, interactive ELF. It covers a wide range of speech events in Asian settings, from press conferences and service encounters to seminars and conversations. All three corpora are available for research purposes. In addition to the publications generated directly by the corpus analyses (see the VOICE, ELFA and ACE websites for respective lists of publications), ELF research can also be found in the dedicated journal, started in 2012, *Journal of English as a Lingua Franca (JELF)*, published twice yearly and currently edited by Barbara Seidlhofer.

In addition to these corpora specifically designed to capture ELF usage, certain other “mixed” corpora include, in varying proportions, spoken and written data by second-
language or EFL users in addition to English L1 texts (e.g. the corpora of academic English MICASE/MICUSP, and BASE/BAWE).

3. The WrELFA corpus

As the previous section makes clear, ELF research to date has focused almost exclusively on spoken English, mainly in interactional settings. Yet for most L2 academics around the world, it is writing and publishing that form a large, or even the major, part of their research activities in English. ELF users have to handle written English at all levels of the publication process and in a wide range of research contexts: not only do they form the great majority of readers of academic publications, they also write articles themselves, act as journal reviewers and editors, book reviewers and editors, or members of PhD thesis boards, submit projects for European or international funding and run international research collaborations, while others engage in scholarly blogging and more informal written exchanges. And the stakes are high: both individual careers and the advancement of learning hinge on the quality of the research carried out and on the coherence and strength of the argumentation, which are primarily expressed through the language used. It is these sorts of considerations that motivated the ELFA team in 2011 to start collecting the very first written ELF corpus, comprising three academic genres: unedited research articles, PhD examiner reports, and research blogs. The 1.5 million-word corpus was completed in April 2015 (see Table 1). The latter two genres were selected for two main reasons:

In the first place, they don’t pass any linguistic gatekeepers or language experts between the author and the intended audience. Secondly, the texts are written by active scientists ranging from PhD students [...] to senior professors [...]. We’re thus in the domain of successful academics using ELF as their means of professional self-representation. (Ray Carey, post on the ELFA Project blog, 29 Sept. 2013)

Similar selection criteria were applied to the research articles subcorpus: the author(s) should not have English as an L1, and the text should not have undergone professional proofreading services or language checking by an English native speaker.

The same general principles were applied for the internal structure and text encoding of the three subcorpora in WrELFA. In each subcorpus, texts are divided into Science (natural sciences, medicine, etc.) and SSH (social and behavioral sciences, humanities). While this is of course not a watertight distinction, it should enable some broad disciplinary comparisons to be made. All the texts were anonymized, regularized to UK spelling conventions, and are available in three formats: .xml, .txt and .rtf. The corpus is not POS-tagged; further details can be found in the corpus manual. The corpus is available on request to any interested researchers.

Table 2 summarizes the full WrELFA corpus and a brief description of the contents of each of the three sub-corpora is given below; a more detailed breakdown of the corpus composition can be found on the ELFA website.

Table 2. Composition of the WrELFA corpus

<table>
<thead>
<tr>
<th>Sub-corpus</th>
<th>N° of texts</th>
<th>N° of words</th>
<th>% by discipline</th>
<th>Ave. words per text</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.1. Research blogs subcorpus

The smallest of the three components in terms of word count is the research blog subcorpus. With 372k words, it represents 24% of the complete WrELFA corpus. It consists of a sample of posts and comments from forty research blogs, tracked down via the aggregator researchblogging.org, and all maintained by L2 users of English. Among the commenters, who account for less than a quarter of the total word count, there may of course be some L1 English speakers, though their precise percentage is difficult to determine. A majority of the blogs (23/40) and words (59%) are by PhD students and junior academic staff, while only eight blogs are maintained by senior academics. The blogs subcorpus is heavily skewed towards the natural sciences, with thirty-three blogs, against only seven for the social sciences and humanities. This reflects a general trend in research blogging, an activity that is more prevalent in the sciences.

3.2. PhD examiner reports subcorpus

This slightly larger subcorpus (402k words, representing 26% of the total) is the most “local” in origin, as all the reports were collected from the University of Helsinki, in 2011–12. The reports are preliminary reports written by the external examiners (rapporteurs in the French context) before the thesis defense; the post-defense reports (rapports de thèse) were not included. Among the 531 reports written in English during the collection period, only 330 authors (63%) gave their permission for the report to be included in the WrELFA corpus, indicating the difficulty of compiling a corpus of this hitherto unexplored “occluded genre” (Swales 1996). In contrast to the blogs subcorpus, the PhD reports are fairly evenly distributed between science and SSH (46% and 54% respectively), and unsurprisingly are mainly authored by senior academic staff. This subcorpus is however the one with the largest proportion of L1 English authors, who account for 29% of the word count – who are albeit writing for an L2 readership –
followed by L1 Finnish (15%) and German (13%) authors. Altogether thirty-three different L1s are represented.

3.3. SciELF subcorpus

This is the largest subcorpus, with 759k words, accounting for 50% of the total word count in WrELFA, and also the latest addition. It was compiled thanks to an international network of researchers in academic discourse from ten different countries: Finland, Czech Republic, France, China, Spain, Russia, Sweden, Italy, Brazil and Romania. We were contacted in autumn 2013 by Anna Mauranen, asking for our collaboration in collecting unedited final drafts of RAs by researchers in our respective L1 countries. The initial target was for each partner to contribute twelve articles, each roughly 7,000 words long, six in Science and six in SSH, bringing the total per country to between 84k and 100k, and the complete corpus to over a million words. As some partners quickly pointed out, many RAs in science are much shorter than 7k words of running text, and this difference with SSH is reflected in the final composition of the corpus (Table 2): despite the slightly larger number of Science articles (78 vs 72 in SSH), the total number and percentage of words are less than those in SSH, due to the lower average number of words per article (4,185 in Science vs 6,012 in SSH).

The methods of collection adopted by the partners in the ten countries probably differed, depending on local circumstances. While the project leaders in Finland were able to call on the Language Services at the University of Helsinki, which conduct professional language revision for university faculty, in helping them to locate suitable material, most partners relied in all likelihood on personal contacts and/or stocks of articles they had proofread or edited for colleagues. This was the case for the French contribution: I contacted sixteen researchers (nine in science and seven in SSH) whose articles I had corrected in the last couple of years and that corresponded to the corpus selection criteria. As for all the other partners’ contributions, the request was accompanied by a description of the project aims and an agreement to sign giving permission for their article to be included in the database. They all unreservedly gave their permission. As can be seen in Table 3, while almost all partners met the initial target figure for the number of articles, the word counts were sometimes well below the 84k mark. As a result, some L1s account for a greater proportion of the total than others, resulting in a slight imbalance. The French contribution is the third largest in terms of word count.

A wide range of investigations is possible using SciELF: interdisciplinary or interlinguistic (author L1) differences, detection of emerging trends specific to ELF academic English, etc., all of which should provide relevant input for ESP trainers and discourse analysts. As a first step, a small-scale analysis of how enabling verbs (allow, enable, permit) are used in SciELF is proposed in the following section.

### Table 3. The ten L1 categories in the SciELF corpus

<table>
<thead>
<tr>
<th>First author’s L1</th>
<th>No. of articles</th>
<th>No. of words</th>
<th>% of words</th>
<th>avg. words/article</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Finland</td>
<td>25</td>
<td>123,153</td>
<td>16%</td>
<td>4,926</td>
</tr>
<tr>
<td>2 Czech</td>
<td>22</td>
<td>109,173</td>
<td>14%</td>
<td>4,962</td>
</tr>
<tr>
<td></td>
<td>Language</td>
<td>Number</td>
<td>Total</td>
<td>Percentage</td>
</tr>
<tr>
<td>---</td>
<td>-----------------</td>
<td>--------</td>
<td>-------</td>
<td>------------</td>
</tr>
<tr>
<td>3</td>
<td>French</td>
<td>16</td>
<td>91,186</td>
<td>12%</td>
</tr>
<tr>
<td>4</td>
<td>Chinese</td>
<td>21</td>
<td>84,807</td>
<td>11%</td>
</tr>
<tr>
<td>5</td>
<td>Spanish</td>
<td>13</td>
<td>79,038</td>
<td>10%</td>
</tr>
<tr>
<td>6</td>
<td>Russian</td>
<td>13</td>
<td>71,376</td>
<td>9%</td>
</tr>
<tr>
<td>7</td>
<td>Swedish</td>
<td>13</td>
<td>60,060</td>
<td>8%</td>
</tr>
<tr>
<td>8</td>
<td>Italian</td>
<td>11</td>
<td>58,685</td>
<td>8%</td>
</tr>
<tr>
<td>9</td>
<td>Portuguese (Brazil)</td>
<td>12</td>
<td>56,625</td>
<td>7%</td>
</tr>
<tr>
<td>10</td>
<td>Romanian</td>
<td>4</td>
<td>25,197</td>
<td>3%</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>150</td>
<td>759,300</td>
<td>100%</td>
</tr>
</tbody>
</table>

4. Enabling verbs in SciELF

Enablement, defined as the provision of adequate power, means, opportunity or authority to do something (Huddleston & Pullum 2002), is a key parameter in research activities, as the research program can only be pursued if circumstances allow it: availability of sufficient data, equipment, methods, or funding; a body of theory to understand and make plausible interpretations of results; sufficiently conclusive results to be able to put forward a claim, etc. Confirmation of the frequency of this lexical category in academic writing can be found in Biber et al. (1999: 713), who found that the register with the highest number of occurrences of the active forms of allow and enable is the ACAD (i.e. Academic) register (with respectively over 100 and 50 occurrences per million words). It therefore seemed relevant to investigate how enabling structures were used in ELF academic writing, using the SciELF corpus.

Three verbs were selected, namely allow, enable, and permit (V1), followed by a second verb (V2). The standard pattern is:

V1 (allow/enable/permit) + NP + to-infinitive

Only the active voice was considered (1), excluding passive forms (2) and dependent relative clauses (3):

(1) Our comprehensive dataset enables us to test these associations (SciELF)
(2) The samples were then allowed to cool down to room temperature. (SciELF)
(3) ... the plaintiff, whom the coach had permitted to use the gym, ... (BNC)

Before analyzing the SciELF corpus, a search was done for the standard pattern in the written Academic sub-corpus of the British National Corpus, to check whether native-speaker usage conformed to the pattern. For permit and enable, no exceptions to the rule were found. With allow, there were nine cases of missing NPs. Four occurred in experimental protocols (4); this type of ellipsis, in which the NP is recoverable from the co-text, is very common in procedural texts such as cookery recipes and protocols:

(4) Drain the grid by touching the edge to filter paper, allow to dry partially under an incandescent lamp (BNC)
The remaining five were all from the same two scientific articles, which on closer inspection turned out to have been authored by Italian and Spanish researchers; they apparently slipped through the net when the BNC was being compiled (5):

(5) The gel retardation assay is a rapid method which **allows to compare** the affinity (BNC)

It can therefore be considered that the standard pattern is a fully respected norm in NS usage.

The SciELF corpus was then searched using the concordancer AntConc (3.4.4) and occurrences were manually examined to detect any non-canonical uses by the ELF authors with respect to the standard pattern. Table 4 gives an overview of the results.

**Table 4. Deviations from the standard pattern of active enabling verbs in the SciELF corpus**

<table>
<thead>
<tr>
<th></th>
<th>Total occurrences</th>
<th>Deviations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science</td>
<td></td>
<td></td>
</tr>
<tr>
<td>permit</td>
<td>26</td>
<td>25</td>
</tr>
<tr>
<td>enable</td>
<td>21</td>
<td>16</td>
</tr>
<tr>
<td>allow</td>
<td>77</td>
<td>41</td>
</tr>
<tr>
<td>Total Science</td>
<td>124</td>
<td>82 (70%)</td>
</tr>
<tr>
<td>SSH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>permit</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>enable</td>
<td>47</td>
<td>3</td>
</tr>
<tr>
<td>allow</td>
<td>98</td>
<td>30</td>
</tr>
<tr>
<td>Total SSH</td>
<td>147</td>
<td>35 (30%)</td>
</tr>
<tr>
<td>Total Science + SSH</td>
<td>271</td>
<td>117 (43%)</td>
</tr>
</tbody>
</table>

Out of the total 271 occurrences of the pattern that was studied, almost half (117, or 43%) were deviations, or non-canonical uses, though with a marked imbalance between science, which accounted for 70% of the non-canonical structures vs 30% in SSH articles, even though the structure was used slightly more frequently in SSH than in science (147 vs 124 occ.). The vast majority of deviations (107 out of the total 117, or 92%) concerned two structures: first, omission of the NP between V1 and V2 (examples 6); this was the most frequent, accounting for two-thirds of the deviations; and second, use of V-ing instead of the to-infinitive in the V2 slot, again with omission of the NP (examples 7):

(6) This analysis allows to understand the communication strategies (SSH) improved precision of CRE dating has **enabled to verify** ages of major failures (Sci)
It was proven that O, permits to remove compounds present on AC surface (Sci)

(7) this hypothesis does not allow explaining all the results (SSH)
This linkage has **enabled establishing** bridges (Sci)
The few other incorrect structures (ten cases, or 8%) were omission of to before the infinitive (allows obtain), or idiosyncratic structural combinations (e.g. will allow to automatically identifying).

In view of the high rate of deviations among ELF writers (43%), this pattern shows a clear difference with NS usage, particularly among science authors, though at this stage in the corpus analysis, the reasons for the SSH/science imbalance can only be speculated upon. What is very striking, however, is the highly uneven distribution of deviations among the ten L1 categories, as shown in Figure 1, which gives the percentage of the total number of deviations represented by each L1, and, for comparison, the proportion that each nationality represents in the SciELF corpus. Swedish and Chinese authors, who account for 8 and 11% of the corpus respectively, unfailingly use the pattern correctly, accounting for 0% of deviations. Some nationalities (e.g. Finnish, Czech) account for a much lower proportion of deviations than their share in the corpus (4.3 and 7% of deviations vs 16 and 14% of the corpus respectively), whereas for others (Portuguese, Italian, Russian, Spanish) the percentage approximates their share of the corpus. The most striking difference is the French L1 authors, who have the dubious distinction of representing 12% of the corpus data but 54% of deviations in this pattern.

5. Interest of ELF corpus data for ESP

For reasons of space, only one type of structure has been briefly examined here, but many other types of queries on the corpus – lexis, syntax, discourse/argumentation structure, citation practices, etc. – can, of course, be conducted. From a contrastive linguistics perspective, this type of L1-specific information can clearly lead to a variety of
comparative studies that should enrich our knowledge of genre variation in RAs, and show whether the master genre is appropriated in different ways by different language groups. The two broad sub-corpora of science and the humanities can also be approached from a contrastive perspective.

28 From an EFL perspective, the deviations such as those highlighted above would be regarded as errors, probably attributed to a carryover effect from the author’s L1 – e.g. for French, the influence of permettre de faire in cases such as example (6) – and requiring remedial action by trainers or correction by language editors of manuscripts for publication. The SciELF corpus provides an extremely useful database for ESP trainers in academic writing, whether they are dealing with mono-language (e.g. all French-speaking) or multi-language groups of graduates, as it will enable them to select the structures most likely to pose problems for their particular students, and propose practice exercises, using concordance lists to highlight the recurrent types of error. As all the occurrences are fully contextualized, solutions can be proposed which take into account aspects that are difficult to teach otherwise, such as information structure considerations, and that highlight the fact that there is not a ‘one-size-fits-all’ solution for many language features. Depending on the context, for example, the function of enablement can be rendered by a wide range of structures and lexical choices: can do, X can be done, is capable of doing, make it possible to, thanks to, omission of the enabling verb altogether, or – when the enabling structure is used, as is frequently the case in French and French L1 writers in English, as a periphrasis for the action itself and not the possibility of action – replacement of the enabling verb by a simple present or past, to mention but a few possibilities.

29 From an ELF perspective, it is interesting, first of all, to note that the non-canonical patterns occur mainly among the Romance languages, which account for 40% of the corpus but 77% of the non-canonical uses; they are therefore unlikely to pose a communication problem (or even be noticed) among ELF users in this group. It could be objected, however, that while this is true of spoken ELF exchanges (including those between academics), in which intelligibility between speakers is the primary concern (Jenkins 2002: 83) and expectations of grammatical or phonological correction often quite low, leading to a high tolerance of deviations from the NS norm, these non-canonical patterns would nonetheless be unacceptable in academic writing. Following this line of reasoning, it would therefore not be possible to define a Lingua Franca Core (see section 2) for academic writing, where only the NS norm should continue to prevail, and the situation cannot be considered on a par with ELF spoken exchanges.

30 However, a cursory search on Google Scholar among published research articles shows that this is not necessarily the case: the non-canonical patterns are finding their way into print in both science and the humanities in top-ranking journals and not just in less prestigious journals or genres that are not peer-reviewed. There are probably several reasons for this: the majority of readers of RAs as well as the editors and reviewers for many major international journals are themselves ELF users and do not pick up these features or see them as an obstacle to comprehension, ELF authors cannot always afford the services of professional language editors for their submissions, journals do not have the resources to do language editing, or – a more radical explanation – academic ELF users are beginning to have an impact on the language of academic writing. Indeed, as shown by Rozycki and Johnson (2013), articles awarded the Best Paper prize in the prestigious IEEE Transactions in engineering contain many examples of non-canonical
This appears to be a more or less spontaneous, grass-roots process that has evolved to meet primarily disciplinary needs in international research communication: as they conclude (2003: 166), “without any conscious proponents, and without any explicit philosophy opposing language imperialism, NNSEs in the field of engineering have organically grown a language that allows all language speakers to communicate with success.” Further and larger-scale investigations in other disciplines, pioneered by the SciELF corpus, should provide us with more data on this potential evolution.

I would like to thank all the French researchers who agreed to contribute their unedited manuscripts to the corpus for their unhesitating cooperation and their interest in the project.

BIBLIOGRAPHY


ELFA website <www.helsinki.fi/englanti/elfa/elfacorp>.


The aim of this research note is to introduce a new corpus that became freely available to researchers in 2016, the WrELFA corpus (Written English as a Lingua Franca in Academic Settings). WrELFA is the written component of the ELFA project directed by Anna Mauranen at the University of Helsinki; the spoken component of ELFA has been available since 2008. WrELFA consists of 1.5 million words drawn from three academic text types: unedited research papers (SciELF corpus), PhD examiner reports, and research blogs. The authors of the texts are academic users of English as a Lingua Franca (ELF), and the texts have not undergone professional proofreading or checking by an English native speaker. As one of the international team of contributors to the SciELF sub-corpus, I will focus on this component of WrELFA, which comprises 150 research articles written in English by authors from ten different L1 backgrounds. To illustrate the interest of SciELF for research and ESP training, a brief analysis of enabling verbs (allow, enable, permit) is presented.

L’objectif de cette note de recherche est de présenter un nouveau corpus qui a été mis librement à la disposition des chercheurs en 2016, le corpus WrELFA (Written English as a Lingua Franca in Academic Settings). WrELFA est la composante écrite du projet ELFA piloté par Anna Mauranen à l’Université de Helsinki; le corpus oral du projet est disponible depuis 2008. WrELFA comprend 1,5 million de mots dans trois genres discursifs universitaires : articles de recherche (corpus...
SciELF), rapports de thèse, et blogs de chercheurs. Aucun texte n’a été revu ou corrigé par un anglophone ou un correcteur professionnel, et tous les auteurs sont des utilisateurs ELF (anglais comme lingua franca). En tant que membre de l’équipe internationale qui a recueilli des articles du corpus SciELF, je me focalise surtout sur cette composante, qui comprend 150 articles écrits en anglais par des chercheurs de dix L1 différentes. Pour illustrer l’intérêt de SciELF à des fins de recherche et de formation en anglais de spécialité, une première analyse des verbes de permission (allow, enable, permit) est présentée.

INDEX

Mots-clés: article de recherche, corpus SciELF, écrit universitaire, ELF, anglais comme lingua franca, verbe de permission

Keywords: academic writing, ELF, enabling verb, English as a lingua franca, research article, SciELF corpus

AUTHOR

ELIZABETH ROWLEY-JOLIVET

Elizabeth Rowley-Jolivet is a research member of the Laboratoire Ligérien de Linguistique at the University of Orleans, France. Her research interests cover multimodality in spoken and written scientific discourse, the epistemology of science, genre analysis of academic discourse, web-mediated genres, and English as a Lingua Franca in academic settings. She has published in several international journals, compiled or contributed to several corpora in ESP and ELF, and co-edited books in these areas. elizabeth.jolivet@univ-orleans.fr