Leah Gilner*
Identification of a dominant vocabulary in ELF interactions

共通言語としての英語 (ELF) の優勢語彙の検証

Abstract: This paper reports on several studies whose common theme is the elicitation of the lexical preferences of speakers of English in localized and globalized settings. Findings from analyses of various corpora show that there exists a relatively small set of preferred words that speakers of English rely on regardless of where the interaction takes place, with whom they are interacting, and what the purpose of the interaction is. Results also show that these lexical preferences are consistently prevalent to the extent that it is possible to advance the hypothesis that a relatively stable dominant vocabulary dynamically emerges out of ELF speaker interactions in order to serve certain communicative functions.

Keywords: ELF, vocabulary use, dominant vocabulary, communication strategies, VOICE, ELFA, ICE

要旨: 本論文は、英語使用者の優勢語彙を、 コーパス分析に基づいて調査した複数の研究についての報告である。各研究は、優勢語彙を見分けることを共通の目的とし、グローバル及び特定地域の双方の状況において用いられる言語が果たす役割を検証することに焦点を当てている。これらコーパス分析からのさまざまな研究結果は、どのような場面で、誰と、どのような目的のやり取りであるかに関わらず、英語使用者は比較的少数の高頻度語彙に依存するところが大きいことを見出している。これらの研究結果から、共通言語としての英語を使用する話者のやり取りから比較的安定した優勢語彙がダイナミックに出現するという仮説を更に深めることが可能であると考えられる。また、話者はコミュニケーションの結合価に基づき、これらの言葉を好むと推定できる。

キーワード: 共通言語としての英語, 語彙使用, 優勢語彙, コミュニケーション方略, VOICE, ELFA, ICE

*Corresponding author: Leah Gilner, Bunkyo Gakuin University, Bunkyo, Tokyo 113-8668, Japan, E-mail: gilner@bgu.ac.jp

1 Introduction

The description and documentation of the English language has traditionally been constrained by a narrow approach that has framed the discussion within conventions and behaviors claimed to be exhibited, and solely exhibited, by typically monolingual, mother tongue speakers. The rapid and extensive diversification of speakers who use English to communicate with one another has effectively overwhelmed established frameworks. We are in the midst of an active and exciting period of reconceptualization and speculation that will inevitably lead to a more realistic configuration of the problem space of linguistic enquiry. There is much to be done, in particular much to be described, before it becomes possible to formalize our understanding of contemporary English language use.

The relevance of corpora in this descriptive undertaking cannot be over-emphasized. "Corpora have been used extensively in nearly all branches of linguistics including, for example, lexicographic and lexical studies, grammatical studies, language variation studies, contrastive and translation studies, diachronic studies, semantics, pragmatics, stylistics, sociolinguistics, discourse analysis, forensic linguistics, and language pedagogy" (McEnery and Xiao 2005: 161). This is so because the intended representativeness of a principled collection of language samples, a corpus, provides a vast yet finite universe of linguistic artifacts upon which it is possible to perform formal inspections within a set of known assumptions (i.e., the principles of compilation). The large-scale nature of these bounded universes of communicative events provides ample grounds upon which to pursue the isolation of emergent phenomena and their classification into stable patterns otherwise too subtle or too complex to discern. The studies discussed in this paper make use of these investigative tools to discern lexical preferences from the composite of vocabulary choices made by English speakers when they interact with one another.

The study of global English language use, in general, and of ELF interactions, in particular, is supported by the availability of the following corpora, all of which will be discussed in this paper. The International Corpus of English (ICE) provides representative samples of Inner and Outer-circle varieties of English with "[...] the primary aim of collecting material for comparative studies of English worldwide" (Nelson 2011). The 26 English varieties corpus (26EV) provides a comprehensive representation of written colingual English language use around the world. The Corpus of English as a Lingua Franca in Academic Settings (ELFA) and the Vienna-Oxford International Corpus of English (VOICE) provide complementary collections of spoken English as a lingua franca (ELF).

It is relevant to note that the discussion will be framed within prevailing perspectives relating to the communicative strategies employed by English speakers who possess diverse linguacultural resources. It will be argued that a direct consequence of this diversity is the enhanced role of negotiation and that one result of this negotiation is the emergence and exploitation of a shared dominant vocabulary. Furthermore, it will be posited that the use of this dominant vocabulary lends support to the observation made by Jenkins et al. (2011) regarding how speakers signal and construct identity while at the same time ensuring intelligibility. The emergence of these lexical preferences could also find interpretation within the notions advanced by Seidlhofer (2009) concerning the ways in which interlocutors promote cooperation and signal territoriality while working together to define their own interactional space.

ELF has been defined as "any use of English among speakers of different first languages for whom English is the communicative medium of choice, and often the only option" (Seidlhofer 2011: 7). The very definition implies diversity among interlocutors. As such, "[t]raditional notions of language or speech community are somewhat unhelpful for setting descriptive parameters for ELF as a social phenomenon" because they do not take the transient, mobile, multilingual nature of the users into account (Mauranen 2012: 23). The construct of linguistic repertoire seems more befitting since it acknowledges that individuals embody a personalized repertoire of linguacultural resources and preferred communicative devices. Blommaert and Rampton (2011: 4f) explain that these linguistic repertoires vary greatly across individuals due to the "differentially shared styles, registers and genres, which are picked up (and maybe then partially forgotten) within biographical trajectories that develop in actual histories and topographies." A central question in ELF study concerns how individuals with widely varied, diverse, and typically multilingual repertoires manage to converge on mutually satisfactory means by which they can accomplish their communicative ends. "In ELF interaction, the interlocutors cannot depend on shared linguacultural conventions and so they have to find common ground by developing their own local conventions in flight as it were, as appropriate to their own contexts and purposes" (Widdowson 2015: 366).

Examination of ELF interactions shows that ELF users are active agents who monitor and adjust to their interlocutors (Firth 2009; Jenkins et al. 2011; Mauranen 2012), engaging in the "strategic negotiation of the linguistic resources" that facilitate "the co-construction of understanding" (Seidlhofer 2011: 198). Successful negotiation implies consensus from all parties on the rules and tools of engagement, one of which is proposed to be a shared repertoire of linguacultural resources. Shared repertoires can be described as emergent linguistic and communicative practices that a given group of interlocutors converges upon to foster

mutual understanding and to create an affective interactional space. Research findings indicate that successful ELF users do not confine themselves to particular linguistic forms but rather demonstrate a willingness and ability to exploit their linguacultural resources in ways that will get the meaning across in a particular moment to particular participants (Mauranen 2012; Seidlhofer 2011). "It is thus not so much uniformity of form, but communicative alignment, adaptation, local accommodation and attunement that would appear to underpin successful lingua franca interactions" (Firth 2009: 163).

The hypothesis explored in this paper is that lexical distributions of spoken ELF interactions identify high frequency words as a reliable resource that serves the function of a "globally valid solution" (Mauranen 2012: 32) which can be used to fill anticipated gaps in assumed shared experience. My contention is that the high frequency words are an integral part of individuals' linguistic/language repertoires and that, through experience with and exposure to other speakers in varied communicative events, these words become preferred communicative assets because of the high probability that a given conversational partner will also make use of them. In Mauranen's (2012: 32) words: "Safe guesses in respect to what an interlocutor might find comprehensible are likely to be features that are widely shared among the world's English speakers."

2 The corpora under consideration

This paper reviews a series of studies into the lexical distributions of vocabulary choices of English speakers in localized and globalized settings with particular interest in a dominant vocabulary composed of high frequency words in and across these corpora. Specifically, this paper makes use of results from analyses published elsewhere of ICE (Gilner and Morales 2011; Gilner 2015), 26EV (Gilner et al. 2011, Gilner et al. 2012; Gilner 2015), VOICE (Gilner 2014a, Gilner 2015), and ELFA (Gilner 2015).The particulars relevant to each specific analysis are given in the corresponding publications. Suffice it to say for our present purposes that these studies document an expanding investigative trajectory into the role of high frequency words in English language use. Over time, the scope of enquiry has broadened, in accord with perspectives in the field, to encompass world Englishes and ELF. Cumulative findings indicate that English-language users around the world rely heavily on a relatively small set of words to communicate in both speech and writing, for private and public purposes, in intra- and international settings. A general description of each of the corpora follows.

The International Corpus of English (ICE) project was started in order to provide a systematic means to conduct comparative studies of Inner and Outer

Circle varieties of World Englishes (Nelson 2011). In order to ensure compatibility among individual corpora, ICE enforces certain guidelines. Specifically, each corpus contains 500 texts of approximately 2,000 words each, collected from 1990 on; approximately 60 % of a given corpus reflects spoken discourse represented by 100 private and 80 public dialogs as well as 120 scripted and unscripted monologues, while approximately 40 % of the samples capture written discourse in the form of 30 letters and 20 student writings along with 150 printed texts originating in instructional, academic, literary, newspaper, and other domains. Speakers/writers are both male and female, 18 years old or older, and educated in the respective country. In this manner, ICE provides a means of analyzing spoken and written discourse of a particular variety (Nelson 2011). The following seven ICE corpora were used for the studies reported here as they were freely available at the time of analysis: Canada, East Africa, Hong Kong, India, Jamaica, the Philippines, and Singapore.

The 26 English varieties (26EV) corpus was the result of a centralized, coordinated, and relatively inexpensive gathering of text samples carried out by the author and colleagues (Gilner et al. 2011). The corpus amounts to 7,800 written texts and approximately 15 million words organized into 26 sub-collections of 300 texts each. The compilation process followed the principles outlined by Sinclair (2004); in particular, every effort was made to represent a balanced view of each variety in terms of mode, type, domain, language, location, and date. Furthermore, the contents were selected according to their communicative function in the community in which they arose rather than for the language they contained. The corpus accounts for three types of discourse for each of the varieties under investigation, specifically, government documents (Hansards, court rulings, legislation), newspaper articles, and opinion columns. The English varieties sampled were: Australia, the Bahamas, Belize, Bermuda, Cameroon, Canada, Fiji, India, Ireland, Jamaica, Kenya, Liberia, Malawi, Malaysia, Myanmar, New Zealand, Nigeria, Pakistan, the Philippines, Singapore, South Africa, Sri Lanka, Trinidad and Tobago, the United Kingdom, the United States, and Uganda.

The Vienna-Oxford International Corpus of English (VOICE) was the first freely available electronic compilation of data seeking to provide an empirical basis for the analysis of spoken English as a lingua franca. The corpus contains transcripts of naturally occurring, non-scripted face-to-face interactions. It amounts to about 1 million words of spoken ELF, corresponding to approximately 120 hours of transcribed speech. The speakers recorded for VOICE typically are experienced ELF users from a wide range of first language backgrounds. VOICE includes samples from 752 individuals, mainly from European countries, with approximately 50 different first languages (Corpus

Description 2013). The corpus data have been categorized into domains (professional, educational, leisure) as well as into speech event types, all of them polylogic: conversation, interview, meeting, panel, press conference, question–answer session, seminar discussion, service encounter, working group discussion, workshop discussion.

The Corpus of English as a Lingua Franca in Academic Settings (ELFA) was created to meet the need for a means of investigating English as it is used by the international academic community to discuss, disseminate, and exchange knowledge, findings, and criticism worldwide (ELFA 2008). The corpus contains transcriptions of approximately 131 hours of naturally occurring academic ELF in both monologic and dialogic speech events drawn from English-medium instruction graduate courses in Finnish universities as well as professional academic ELF as used in seminars, conferences, and doctoral defenses (Carey 2013). Speakers come from more than 50 countries and are described as well educated, plurilingual, and of variable proficiency. ELFA amounts to slightly over 1 million transcribed words. It is structured according to disciplinary domain as well as according to speech event type. Speech event types are further subdivided according to setting (conference, doctoral defense, lecture, panel, and seminar) as well as clustered into monologic and polylogic events.

3 Lexical distributions of high frequency words

Although there are many different reasons why English speakers choose to use certain words more often than others, corpus analysis makes it possible to reliably identify a set of words that are frequently used regardless of mode, speaker, genre, or domain. The data in the following two tables were prepared for Gilner (2014b), a presentation given at the Waseda ELF International Workshop 2014. The results reproduced here were obtained by conducting straight frequency counts of unlemmatized words in ICE, VOICE, and ELFA. Please note that where these numbers differ from those given elsewhere by other researchers, the differences are likely due to the treatment of fillers and other non-words by different methodologies (here, for example, removed).

Table 1 presents the most preferred words found in ELFA, VOICE, and the spoken component of ICE together with a tally of their repeated occurrences grouped into frequency bands.

Lexical distributions in English are extremely sloped. Very few words account for the majority of lexis used in speech and writing. Those unfamiliar with lexical distributions are generally surprised to discover that the 100 most

Table 1: The most preferred vocabulary in ELFA, VOICE, and ICE spoken.

Top words	VOICE (%)	ELFA (%)	ICE spoken (%)
100	61.57	57.20	54.40
200	69.61	64.63	62.68
500	78.03	73.20	71.85
1,000	83.36	79.27	78.17
1,500	86.11	82.56	81.54
2,000	87.79	84.69	83.82

frequent/preferred words and their repetitions account for more than half of any given stretch of discourse. The findings presented in Table 1 exemplify this characteristic of English language use. The presence of a dominant vocabulary is clearly evident in all three corpora.

Specifically in the case of VOICE, an inspection of these distributions yields that the 100 most preferred words account for 61.57 % of the entire corpus. That is, if VOICE had exactly 1 million words, 615,700 of them would be repetitions of these 100 most preferred words. It is important to notice that, in contrast, the second 100 most frequent/preferred words account for only about 8 % of VOICE. That is to say, speakers choose to use the first 100 most frequent words much more often than the second 100 most frequent words. This decreasing trend of occurrence continues as the degree of preference diminishes. The 500 words that separate the 1,500 and the 2,000 frequency band account for only 1.68 % of all the running words in VOICE. This trend holds across corpora.

The cumulative preference for this specific vocabulary is remarkable. The 2,000 most frequent/preferred words account for 87.79 % of VOICE. Performing the same frequency analysis on ELFA shows that the numbers observed in VOICE are not idiosyncratic but rather appear to be an emergent feature of language use in general. The frequency bands for ELFA show that the most preferred vocabulary is still highly dominant although slightly less so than in the case of VOICE, an observation that will be addressed in more detail later on. The 2,000 most frequent words account for 84.69 % of the entire ELFA corpus. That is, 2,000 words alone account for eight to nine out of every ten words, an observation that also holds true for the data shown for the spoken component of ICE. Therefore, it is important to acknowledge that the most preferred vocabulary is relatively small in number while strongly prevalent in occurrence. We can observe not only that there exists a specific vocabulary that speakers prefer to use but also that speakers' choices make this specific vocabulary a strikingly dominant feature of discourse.

The previous analyses by this author extracted a list of the most frequent or preferred vocabulary from each corpus. It is reasonable to ask if these corpus-specific word lists contain the same vocabulary and, if they differ, how they do so. Table 2 shows the ten most frequent words in VOICE, ELFA, and ICE which, following the same pattern of increasing occurrence according to increasing preference, account for 20 % to 25 % of the running words in each corpus.

Table 2: The top ten most preferred words in VOICE, ELFA, and ICE spoken.

Ranking		VOICE		ELFA		ICE spoken
1st		the		the		the
2nd		and		and		to
3rd		I		of		and
4th		to		to		of
5th		it		in*		I
6th		you		that		you
7th		that		I		that
8th		is*		it		a*
9th		of		you		in*
10th		we*		is*		it
	12th	in*	12th	a*	12th	is*
	13th	a*	16th	we*	14th	we*

The first observation is that these lists are remarkably similar in content. The second observation is that words are often ranked differently across lists and corpora. For instance, the word *in* ranks 5th in ELFA and 9th in the spoken component of ICE. Yet, while missing from the top ten in VOICE, it is nonetheless found in 12th position. Inspecting the top 100 most preferred words across corpora reveals increasingly wider variations in rank. These variations become broader as the frequency of occurrence decreases so that a word like *expect* has a ranking of 597 in VOICE, 1,079 in ELFA, and 865 in the spoken component of ICE. These variations mean that a naïve comparison of the top 2,000 words across corpora would give the false impression of missing words that might just be found over an arbitrary cutoff boundary. The identification of the preferred vocabulary across corpora, if at all possible, requires a more sophisticated approach than a simple comparison of word lists.

4 Identification of a set of shared high frequency words in localized settings

The observation that different corpora yield similar yet superficially different lists of preferred words is one of the reasons that motivated the construction of the word list called the ICE-CORE (Gilner and Morales 2011).

The ICE-CORE word list originates from the analysis and comparison, on equal footing, of the lexical distributions in the seven corpora from ICE previously mentioned. Briefly, the methodology used to identify the ICE-CORE was as follows. First, the lexical distributions for each corpus were calculated, producing a variety-specific frequency list for each. Second, each variety-specific list was used to assess its corresponding corpus. Third, through a number of iterations, the ICE-CORE word list came to contain a relatively sophisticated intersection of these variety-specific lists. A full description of the methodology can be found in Gilner and Morales (2011). In its original form, the ICE-CORE is composed of 1,206 word families.

The organizing principle of word family is a conceptualization of lexical structure and morphological knowledge that dates back to the work of Harold Palmer in the 1930s (e.g., Palmer 1931). Ongoing research supports the thesis that this conceptualization has representational and pedagogical value and that, in particular, awareness of word family relationships can greatly decrease the learning burden of derived words containing known base forms (Nation 2001). Simply put, a word family is a collection of words related through affixation to a base or root form. The base or root form is termed the *headword*. Related words fall into two categories. The first category is composed of words related to the base form through inflectional suffixes. The second category is composed of those words related to the base form through derivational affixation. Bauer and Nation (1993) propose seven levels of affixation. The ICE-CORE word families include words up to level six, that is, those that are formed by orthographically regular affixes as well as some irregular but frequent ones. For example, the word family referred to as *access* is composed of the headword *access*, the inflections *accesses, accessing, accessed*, and the derivations *accessible, inaccessible, accessibility, inaccessibility*.

After the analyses provided in Gilner and Morales (2011), Gilner et al. (2012), and Gilner (2014a) were published, revisions were made to the ICE-CORE. These revisions are justified and documented in Gilner (2015) together with updated results of the original studies. These revisions and updates do not alter the validity of the original papers and those papers remain the source to call upon for reference. The reader is advised that the data reported henceforth are reproduced from Gilner (2015).

5 Use of high frequency words in localized settings, part I

Tables 1 and 2 show that high frequency words account for most of the running words in texts and that this observation holds across domains. The ICE-CORE has been obtained from the high frequency words found in all seven ICE corpora. Gilner and Morales (2011) go on to show that the ICE-CORE is nearly as representative as variety-specific high frequency word lists, thereby encouraging the hypothesis that a shared preferred vocabulary exists across English varieties and that the ICE-CORE closely resembles it. Table 3 presents data that support this proposition.

Table 3: Variety-specific coverage versus ICE-CORE coverage of seven ICE corpora.

	Canada (%)	East Africa (%)	Hong Kong (%)	India (%)	Jamaica (%)	Philippines (%)	Singapore (%)
Variety specific	82.74	82.56	84.85	81.12	83.07	81.09	82.85
ICE-CORE	80.77	81.52	81.45	79.59	81.43	79.54	81.49
Difference	1.97	1.04	3.41	1.53	1.64	1.55	1.36

It is important at this point to draw attention to the term *coverage* as it will be used throughout the paper. *Coverage* refers to the occurrence of a word or group of words in terms of a percentile share of the total running words in a given text sample. If a group is of preferred words, its coverage will be high. If a group is of less preferred words, its coverage will be low, indicating that these less preferred words are used infrequently and, consequently, analyses of lexical distributions will find few occurrences of them in a given text sample.

The results shown in Table 3 were obtained by analyzing each of the ICE corpora with its corresponding variety-specific list of preferred words and the ICE-CORE list. For consistency, all word lists contained the same number of lemmatized words (no derivations were included). Results indicate that variety-specific word lists provide better coverage than the ICE-CORE. This is unsurprising. The words preferred by, for example, Canadians are best represented by a word list extracted from (and only from) a corpus of samples that represent the speech of Canadians. The data show that the same holds true for each corpus and, consequently, for each speech community investigated.

Interestingly, the ICE-CORE provides very similar coverage despite being a variety-neutral word list. The row labeled *Difference* shows a minimum

difference of 1.04 % in the case of East Africa English and a maximum difference of 3.41 % in the case of Hong Kong English. The average difference is 1.79 %, an indication that the ICE-CORE list is a good candidate to represent the preferred vocabulary in all of these corpora regardless of English variety.

Table 3 provides comparative results of lemmatized word lists (headwords + inflections) in order to show the extent to which the ICE-CORE represents a shared preferred vocabulary across localized varieties. Table 4 provides coverage data for each variety by making use of the word family ICE-CORE list (headwords + inflections + derivations) while breaking down the corpora by discourse mode.

Table 4: ICE-CORE coverage of seven ICE corpora.

	spoken		**spoken + written**	
	Coverage (%)	**Words**	**Coverage (%)**	**Words**
Canada	88.81	642,280	85.49	1,069,974
East Africa	87.26	515,047	85.77	1,407,342
Hong Kong	88.44	969,707	86.22	1,452,303
India	86.89	685,376	84.02	1,121,542
Jamaica	88.95	654,581	86.00	1,065,946
Philippines	86.40	683,729	83.93	1,128,509
Singapore	88.22	665,021	85.73	1,095,896
Combined	87.91	4,815,741	85.35	8,341,512

Since both ELFA and VOICE are spoken-only corpora, data reported in this paper make use of the spoken component of ICE except in the case of Tables 3 and 4, where it is felt it is relevant to present analyses for the seven ICE corpora in their entirety. Thus, Table 4 includes data for the spoken and the written components so as to illustrate the impact of written material on vocabulary preferences. The last row, labeled *Combined*, provides results for all seven varieties combined into a single percentage measure that will be used again in Table 6.

The first observation is that spoken numbers are slightly higher than total numbers. This is uncontroversial. Analyses of corpora of written and spoken language consistently yield lower usage of high frequency words for the former as the manufacture of written language allows for a more deliberate choice of vocabulary (e.g., Biber and Conrad 2009; Nation 2006).

The second observation is that the coverage provided by the complete version of the ICE-CORE word list is significantly better than that of the lemmatized version (compare data in Table 3 against data in Table 4). The inclusion of

derivations in the analyses increased the ICE-CORE coverage of these corpora to nearly 90 %, that is, accounting for eight, nearly nine, words out of every ten words.

In sum, the data presented so far support the notion that the ICE-CORE contains the most significant portion of the vocabulary preferred by speakers of these English varieties.

6 Use of high frequency words in localized settings, part II

Gilner et al. (2012) enlarged the scope of enquiry by creating a dataset that could be used to provisionally complement the ICE project. The aim was to assess the ICE-CORE list against as many varieties of English as possible, including those for which corpora do not yet exist. The rationale was simple. The compilation of corpora is a very expensive process and the field might just have to wait years, perhaps decades, for the compilation of additional corpora to take place. It is easy to imagine the complicated logistics underlying the gathering and organization of the data. The ICE project is relying on the work of twenty different research teams. Twenty years since the project got underway, ten corpora are available, three of which require licensing fees.

The 26EV corpus was devised to circumvent these limitations. Furthermore, it is relevant to note that the 26EV corpus is widely divergent from other corpora and, therefore, provides a stringent testing ground of the hypothesis regarding the existence of a shared preferred vocabulary. Succinctly, each variety is represented by 300 written documents from three distinct domains. The government domain is composed of parliamentary Hansards or, when not available or sufficient in number, court rulings or legislation. The newspaper domain contains reporting articles that, while using less specialized speech, are nonetheless representative of a very specific genre. Last, the opinion domain again reflects a distinct type of discourse and, importantly, text samples were gathered from the Internet together with the entire, unedited comment threads. As a whole, the material in the corpus ranges from highly formalized to highly volatile and, from a lexical point of view, rarified vocabulary is given as much importance as code switching and fleeting Internet coinages.

Table 5 is reproduced from Gilner (2015) and shows the results of the coverage analysis of these 26 English varieties using the ICE-CORE word list. The data illustrate how the ICE-CORE list provides substantial representativeness of the preferred vocabulary used by speakers of English varieties around the world.

Table 5: ICE-CORE coverage of 26 English varieties.

	Coverage (%)	Words		Coverage (%)	Words
Australia	85.73	556,040	Malaysia	79.32	482,740
Bahamas	85.45	553,281	Myanmar	78.80	413,505
Belize	82.95	503,681	New Zealand	86.95	568,080
Bermuda	85.44	551,795	Nigeria	83.58	572,992
Cameroon	79.86	355,386	Pakistan	79.09	461,876
Canada	82.40	614,284	Philippines	78.58	576,299
Fiji	81.69	523,760	Singapore	84.61	588,605
India	79.52	498,808	South Africa	81.88	617,674
Ireland	84.85	594,047	Sri Lanka	81.78	550,973
Jamaica	81.92	509,964	Trinidad and Tobago	85.59	525,680
Kenya	84.31	539,746	Uganda	79.66	470,872
Liberia	81.07	554,544	UK	85.17	648,562
Malawi	82.36	481,456	USA	83.97	666,471
			Combined	82.74	13,981,121

That is, it contains the most significant portion of the vocabulary preferred by speakers of these English varieties. In general, eight words out of every ten words in every sub-collection are included in the ICE-CORE list.

Based on Tables 4 and 5, it is possible to tentatively propose that when speakers/writers communicate within the boundaries of a given local speech community, it is reasonable to expect that the preferred vocabulary they choose to use will be among that included in the ICE-CORE list.

7 Use of high frequency words in globalized settings

It remains to be seen if the ICE-CORE list is representative of the vocabulary preferred by English speakers when they communicate across speech communities, that is, when they participate in ELF interactions. Results in Table 6 include the ICE-CORE coverage metrics obtained for the combined spoken component of the seven ICE corpora previously presented in Table 4 along with those of ELFA and VOICE from Gilner (2015).

Inspection of Table 6 indicates that the ICE-CORE list is indeed representative of the preferred vocabulary in both localized and globalized settings (Seidlhofer 2011: 4). This is demonstrated by the combination of the following two insights. First, if a substantial number of frequent/preferred words from any

Table 6: Comparison of coverage by ICE-CORE of ELFA, VOICE, and ICE spoken-only.

Corpus	Coverage (%)	Running words	Entire corpus
VOICE	92.67	906,507	978,210
ELFA	90.24	917,797	1,017,063
ICE (spoken)	87.91	4,233,517	4,815,741

individual corpus were missing from the ICE-CORE list, the coverage numbers could not possibly be so high. Second, approximately 99 % of the ICE-CORE word families are found in each of the corpora. That is to say, the ICE-CORE list does not contain (or contains only residually) words that are not preferred in the interactions captured in VOICE and ELFA.

Interestingly, speakers in VOICE and ELFA demonstrate a higher preference for words in the ICE-CORE than speakers do in localized settings. The ICE-CORE coverage of the spoken component of ICE shown in Table 6 indicates a comparatively lesser reliance on the ICE-CORE that, nonetheless, accounts for 87.91 % of all running words or almost nine out of every ten. The observation is that ICE-CORE coverage of ICE spoken, whether as a whole (Table 6) or by variety (Table 4), is less than that of ELFA and VOICE. Even considering the spoken component alone, the average coverage of ICE by the ICE-CORE is 2.33 % lower than that of ELFA, and 4.76 % lower than that of VOICE.

Based on the analyses of these corpora, it is clear that, regardless of setting, the word families in the ICE-CORE list equate to a dominant vocabulary in each of these corpora. Over 906,000 out of 978,000 words in VOICE are in the ICE-CORE list. In the case of the seven ICE varieties (see Table 4), 4.2 million out of 4.8 million words are in the ICE-CORE list. Speakers' reliance on the preferred vocabulary is striking regardless of setting. Summing up, the use of this preferred vocabulary again does not appear to be constrained by geographical boundaries, pragmatic function, or interactional context.

Indeed, the weight of the argument is such that it will be useful to informally coin this emergent phenomenon as *dominant vocabulary* (DOVO) as it is clear that the preferred vocabulary does not only imply that it is more frequently chosen but that it is most frequently chosen without regard to the kind of primary considerations (e.g., setting, participant, context) that are known to otherwise alter the characteristics and, at times, very nature of linguistic artifacts. For the purposes of this paper, we propose the ICE-CORE to be a sufficiently faithful instantiation of the DOVO used by English speakers since it has been shown to be composed of the most frequent/preferred words found in all corpora analyzed.

As mentioned, Table 6 shows that coverage numbers are somewhat higher in ELF settings than those found in localized, colingual, non-specialized speech. This could imply that the high degree of usage of the DOVO is not idiosyncratic to the particular discourse of these domains, but rather a strategy employed by ELF speakers. In other words, it could be indicative of efficient deployment of a linguistic resource with high communicative valence.

This is important for several reasons. ELFA is not only an ELF corpus. It is also a corpus of academic language with samples gathered from presentations, seminars, doctoral defenses, and other such typical speech events in academia. VOICE, too, is an ELF corpus yet significantly different from ELFA in that it contains material from speech events such as meetings, interviews, discussions, and lectures taking place in professional, educational, and leisure domains. Even more divergent, the seven English varieties in ICE distinguish themselves from both of these ELF corpora, firstly, by the fact that they contain material from a single bounded linguistic community and, secondly, by their very composition. The structure of the spoken component of every ICE corpus contains interviews, phone calls, face-to-face conversations, classroom lessons, broadcasts, legal cross-examinations, business transactions, and unscripted speeches, among other samples. In short, the corpora under investigation differ from each other in many significant ways. Yet, this diverse conglomeration of speakers manages to successfully communicate across a wide range of domains by devoting approximately 90 % of their lexical choices to a DOVO of relatively small size.

DOVO coverage of the 26EV corpus (Table 5) shows the greatest disparity with ELFA and VOICE (Table 6). The combined value is at 82.74 % while those of ELFA and VOICE are at 90.24 % and 92.67 %, respectively (corresponding to differences of 7.50 % and 9.93 %). In concrete terms, it means that when speakers find themselves in localized, colingual settings, about eight out of every ten words belong to the DOVO. When these speakers find themselves in globalized settings, usage of the DOVO increases to nine words out of every ten words.

These numbers lend support to the hypothesis that speakers in ELF settings may have a comparatively higher preference for a DOVO than speakers in localized settings. However, it is important to note that the differences observed could be influenced by the design of the corpora investigated. Were this to be the case, it would not be an argument against the representative adequacy of these corpora in general, but rather a comment on the utility of these corpora for this specific type of analysis. As the work is preliminary, all options remain open.

These results, combined with those presented previously, prompt certain speculations regarding lexical preferences. First, a DOVO exists in every

localized variety. Second, this DOVO is largely the same across varieties. Differences exist but these are mostly in ranking, not in content. Third, when speakers from diverse linguistic and cultural backgrounds interact in globalized settings, they converge on an ELF-DOVO, drawing on it with slightly but significantly greater emphasis than speakers do when in localized settings. In sum, the ELF-DOVO is moderately more dominant than a given local DOVO such as Jamaican-English-DOVO or Philippine-English-DOVO. This feature-based analysis thus provides discernible data that allow us to partially characterize the lexical preferences of English speakers around the world. It also makes it possible to propose preliminary hypotheses regarding certain aspects of ELF interactions. To this end, and in light of these findings and insights, we will now take a closer look at the ELF data.

8 DOVO in ELF interactions

According to the ELFA corpus compilers, the decision to use broad categories arose from the size of the samples and, in turn, of the corpus because, "[...] otherwise the search results remain too meagre" (Mauranen 2006: 152). In this manner, the breakdown of the corpus according to domain was carried out in terms of high level disciplines. Gilner (2015) provides a DOVO coverage analysis of these domains as reproduced in Table 7.

Table 7: DOVO coverage of ELFA disciplinary domains.

	Coverage (%)	Words	Events	Avg. length
Behavioral sciences	92.36	72,377	10	7,238
Economics and Administration	90.65	53,710	12	4,476
Humanities	91.23	170,261	30	5,675
Medicine	85.68	98,177	17	5,775
Natural sciences	88.81	135,119	15	9,008
Other	93.51	13,064	1	13,064
Social sciences	91.03	279,960	50	5,599
Technology	90.44	194,395	30	6,480
Entire corpus	90.24	1,017,063	165	6,164

Research into why speakers within a particular domain rely more heavily on DOVO than speakers within other domains is beyond the scope of this investigation and, possibly, beyond the scope of ELFA. Nonetheless, it is worth noting that coverage numbers across domains reveal a surprising insight into lexical

preferences that is particularly remarkable considering how highly specialized these domains are: Approximately 1,200 word families dominate lexical distributions no matter the disciplinary field. This is noteworthy because we are instinctively aware of the fact that a doctoral defense in biology is fundamentally different from one in economics. It is unequivocally the case, it is in fact a tautology, that there exists a plethora of linguistic elements that make it impossible to confuse one type of defense with the other and that among these linguistic elements is the choice of vocabulary, the terminology involved. Yet findings show that more than nine out of ten words are repetitions of members of these word families regardless of the domain. From a lexical perspective, speakers are able to express and elaborate with precision upon the radical terminological differences between domains by means of less than one word out of every ten.

Communication is shaped by dynamic processes which influence how speakers choose to express themselves and exploit meaning-making potential. As research findings continue to accumulate, the hybridity, fluidity, and variability of ELF communication is becoming increasingly clear (Cogo 2012). When speakers come together in globalized settings, there appears to be a suspension of assumptions regarding what is "known" among interlocutors (Cogo 2010; Widdowson 2012). As Blommaert and Rampton (2011: 6) explain: "The management of ignorance itself becomes a substantive issue, and inequalities in communicative resources have to be addressed." In order to overcome gaps in shared understanding, participants appear to negotiate and establish norms in real time while orienting discourse toward their listeners (Jenkins et al. 2011; Mauranen 2012; Seidlhofer 2011). Given the findings presented here, it seems that the DOVO is one feature that speakers recognize as shared and that they rely on to promote mutual intelligibility.

It is possible to delve a bit deeper into this insight by examining ELFA and VOICE in terms of broad speech event types. The two umbrella categories, monologic and polylogic, are used to structure the samples according to the number of speakers (i.e., single versus multiple) involved in a given speech event (Mauranen 2006). Table 8, reproduced from Gilner (2015), shows the coverage that the DOVO provides of each category in ELFA.

Table 8: DOVO coverage of ELFA monologic and polylogic categories.

	Coverage (%)	Words	Events	Avg. length
Monologic	87.90	332,401	91	3,653
Polylogic	91.38	684,662	74	9,252
Entire corpus	90.24	1,017,063	165	6,164

Monologic speech event types are mostly or uniquely led by a single speaker, involving longer turns and control of the floor. The use of the DOVO in monologic speech events, while still clearly dominant, is 3.48 % below that of polylogic events. Although polylogic event types account for less than half of all events (44.85 %), they are substantially longer on average, up to 6.74 times. On average, the use of the DOVO in polylogic speech events is 1.14 % over that of the entire corpus.

Thus, the polylogic component of ELFA is comparatively more influential in the corpus as a whole, showing a bias that is acknowledged to be deliberate. Specifically, polylogic speech event types account for 67.31 % of the running words in the ELFA corpus and are represented by discussions in seminars, conferences, panels, and doctoral defenses. According to the ELFA corpus compilers, this decision is based on the insight that "[...] it is in dialogic interaction that language primarily and most naturally gets negotiated" (Mauranen 2006: 153). Indeed, substantial monologues were not included in VOICE for similar reasons: "[...] A central objective of VOICE is to record ELF speech not simply as it is produced by individual speakers but as it happens among speakers in the natural course of interaction" (Breiteneder et al. 2006: 164) in order to capture communicative practices such as negotiation of meaning, back-channelling, feedback mechanisms, clarification devices, etc.

Table 9: DOVO coverage of ELFA speech event types (reproduced from Gilner 2015).

	Coverage (%)	Words	Type	Events	Avg. length
Conference discussion	92.03	72,350	poly	14	5,168
Conference presentation	89.48	93,267	mono	34	2,743
Doctoral defense discussion	90.62	205,155	poly	14	14,654
Doctoral defense presentation	85.07	21,743	mono	10	2,174
Lecture	87.08	139,989	mono	20	6,999
Lecture discussion	91.02	56,588	poly	12	4,716
Panel discussion	93.51	13,064	poly	1	13,064
Seminar discussion	91.69	337,505	poly	33	10,227
Seminar presentation	88.29	77,402	mono	27	2,867
Entire corpus	90.24	1,017,063		165	6,164

Table 9 provides a breakdown of the monologic and polylogic categories by speech event type. Coverage data further specify how discussions (polylogic event types) tend to make comparatively higher use of the DOVO. As logically related pairs, discussions and presentations of conferences are shown to rely the most on the DOVO, while discussions and presentations of doctoral defenses

rely the least. The lowest coverage representative of speaker preferences across ELFA indicates a minimum of 8.5 words out of every 10.

The comparatively higher presence of the DOVO in polylogic speech events could be explained in at least two complementary ways. First, there could be a shift of accommodation strategies in monologic speech events when actual interlocutors are substituted by an imagined one, namely, oneself. The idea is that actual interlocutors offer linguistic material – in this case, words and phrases – that the speaker can reuse once they have been established to be known, thus leading to the recycling of certain vocabulary. In a monologue, and no matter how informative the non-linguistic feedback, the speaker is engaged with an imagined interlocutor (oneself), forcing the speaker to make guesses regarding which vocabulary is shared and known. Accommodation strategies deployed in polylogic speech events are well documented and involve recycling, paraphrasing, repetition, and so on (Cogo 2010; Cogo and Dewey 2006; Jenkins et al. 2011; Mauranen 2012; Seidlhofer 2011). The insight here is that these strategies appear to result in a narrower selection of vocabulary which, in turn, may account for the conspicuous preference for the DOVO shown in Tables 8 and 9. In monologic speech events, these guesses seem to result in a lesser reliance on the DOVO.

Second, monologues might allow the speaker to rely on stretches of rehearsed speech, at least partially, while discussions have a real-time interactional component that forces interlocutors to improvise and produce spontaneous, unscripted discourse. The increased cognitive demands of this latter task naturally compromise online resources and accessibility to less primed lexical items, thus producing fewer vocabulary flourishes. In contrast, prior preparation of material in academic settings is commonplace and one could argue that no worthy formal presentation (the prototypical monologic speech event in ELFA) is ever truly improvised. It can be posited then that premeditation and rehearsals result in a more diverse vocabulary that parallels the comparative increase in lexical diversity observed in written versus spoken discourse (e.g., Biber and Conrad 2009; Nation 2006). If this were the case, the notion of planned versus unplanned discourse could be at play. Georgakopoulou and Goutsos (1997: 34) explain that "[...] writers have time to mould their ideas into a more complex, coherent and integrated whole, making use of complicated lexical and syntactic devices." Hatch (1992: 237) had previously remarked that "[...] as writers revise and polish their performance, the language they use changes" while, in spontaneous speech, "[...] words and phrases are repeated, and words seem to touch off the use of words having similar sound sequences" (Hatch 1992: 241). Furthermore, the argument is that there is a bias toward redundancy in spoken discourse and that repeated words and phrases serve to promote cohesion. The

lexical preferences exhibited in Tables 8 and 9 might be a manifestation of these communicative strategies.

Results from an analysis of the speech event types in VOICE, all polylogic, presented in Table 10 add support to the proposition that interactional factors influence vocabulary choices.

Table 10: DOVO coverage of VOICE speech event types.

	Coverage (%)	Words	Type	Events	Avg. length
Conversation	90.22	149,358	poly	36	4,149
Interview	92.60	34,803	poly	16	2,175
Meeting	92.89	261,178	poly	20	13,059
Panel	91.16	89,804	poly	10	8,980
Press conference	90.55	17,126	poly	5	3,425
Question–answer session	91.70	26,527	poly	10	2,653
Seminar discussion	92.81	61,440	poly	6	10,240
Service encounter	89.71	14,357	poly	11	1,305
Working group discussion	94.41	172,443	poly	19	9,076
Workshop discussion	94.31	151,174	poly	18	8,399
	92.67	978,210		151	6,346

As shown, DOVO coverage of speech event types in VOICE ranges from 89.71 % in the case of *Service encounters* up to 94.41 % for *Working group discussions*. Recall that in Table 9, the highest DOVO coverage of monologic events is lower than the lowest DOVO coverage of polylogic events. Interestingly, this remains true when we look at VOICE, where the lowest DOVO coverage is 89.48 %, still greater than the coverage of the highest monologic type in ELFA. Together, these results give further credence to the notion that the communicative interplay and interchange in polylogic events bear on the degree of utility of the dominant vocabulary.

Another reason why ELF speakers seem to show such marked preferences for the DOVO may stem from the desire to establish rapport. Gómez González (2013), for example, provides an account of the role that lexical cohesion plays in discourse management and turn-taking behaviors. Her analysis documents how interlocutors connect their own speech to their conversation partners' by means of several cohesive devices, the most frequent of which are repetition and associative cohesion. Seidlhofer (2009: 195) observes that ELF speakers co-construct, in real-time, manners and means of expression that "facilitate cooperative convergence on shared meaning." Seidlhofer uses samples from VOICE to illustrate how participants in ELF interactions choose and combine their words

to align themselves with their fellow participants in a pro-tem community of practice. Given the findings presented in this paper, an additional tentative proposition would be that word choice is one way for these speakers to demarcate the group's common ground while at the same time establishing a shared affective space. That is to say, the very words that speakers choose to use may promote a sense of affinity among a given group of interlocutors.

Pitzl et al. (2008: 40) present findings into lexical innovations that illustrate another way that ELF speakers exploit what we here refer to as the DOVO to create "supportive and co-productive" interactional environments. Pitzl et al. (2008) focus on lexical variations found in a small subcorpus of VOICE as identified by the <pvc> tag. The VOICE corpus uses the <pvc> tag for individual lexical items that were not found in the reference dictionary. The researchers approached this preliminary analysis from the perspective of word formation and made observations based on the use of processes such as affixation, borrowing, analogy, and reanalysis. Many of the examples they provide in their paper exploit DOVO as, for example, base forms for affixation. Examples include *increase*, *gather*, *imagine*, *prefer*, and *work*. According to the researchers, the addition of a suffix often served to make meaning more overt and explicit. Mauranen (2006), Ranta (2006), and Seidlhofer (2011) also observe that ELF speakers tend toward communicative strategies (e.g., repetition, rephrasing, and discourse reflexivity) that make meaning more explicit. Other lexical innovations found in VOICE include uses of the prefixes *non-* and *re-* in combination with DOVO words such as *formal*, *read*, *send*, and *confidence*. This tendency was attributed to the economy of expression whereby speakers minimize the number of words needed to express the idea they want to communicate. These findings provide further evidence of how ELF speakers accommodate each other by relying on a shared lexical resource, while at the same time drawing on a sophisticated understanding of word formation potential.

Finally, Carey (2013) discusses results of an analysis of frequent formulaic chunks that further evidences the preference of ELF speakers for what we refer to here as the DOVO. Carey identified the most frequently occurring three- to five-word chunks in the ELFA corpus and found, among other things, that speakers prefer to replace less frequent items with more frequent ones. For example, ELF speakers tend to use *so to say* rather than *so to speak*, replacing the less frequently occurring *speak* with the more frequently occurring *say*. Similar distributions of *so to say* and *so to speak* were observed by Carey in a supplementary analysis of VOICE. The contention is that these strategies further contribute to the comparatively higher use of a DOVO in globalized settings.

9 Conclusion

The findings reviewed in this paper reveal the presence of a DOVO in the discourse of English language users around the world. It has been shown that this DOVO is largely the same across localized English varieties, making it possible to identify a relatively small set of high frequency words that constitutes a shared preferred vocabulary. Comparisons of the presence of this preferred vocabulary in intra-community communication (i. e., localized settings) and inter-community communication (i. e., globalized settings) indicate greater reliance on this shared resource when English is used to bridge linguacultural divides. Furthermore, examination of ELF interactions reveals a higher presence of the DOVO in polylogic than in monologic interactions.

These results support the hypothesis that a relatively stable dominant vocabulary dynamically emerges out of ELF speaker interactions in order to serve the communicative functions of a pro-tem community of practice. It is posited that dominant vocabularies are a component of individuals' language repertories. The continuous evolution of linguistic experiences serves to signal the shared communicative valence of certain words within these vocabularies, equipping the interlocutors with the best lexical candidates with which to engage in strategic negotiation of linguacultural resources of a given group of ELF users. The implication is that, when it comes to dominant lexical preferences, English language users are able to transform a locally established linguistic strategy into a globally valued communicative practice.

This convergence on the DOVO in globalized settings seems to suggest that lexical choice is one way that ELF speakers manage interaction and overcome gaps in shared conventions. Cogo (2010: 296) describes ELF exchanges as those "[...] where people from various backgrounds in more or less stable communities engage in communicative practices that shape, construct and define the communities themselves." Cogo suggests that norms are negotiated by the participants for specific purposes by establishing a shared repertoire of resources that promotes mutual understanding. Findings presented here suggest that the DOVO is one of the resources in this shared repertoire.

The rapid spread of English as a means of international communication requires a shift in theoretical, analytical, and pedagogical approaches to the study of English linguistics and language use. Detailed examinations of how speakers use the language in diverse and expanding interactional situations are necessary if we are to begin to understand the English language as a tool for global communication (Seidlhofer 2011). The results reviewed here come from

investigations which adopted a feature-based descriptive approach with the aim of furthering our understanding of the role of dominant vocabulary in ELF language use. The discussion illustrates how this approach can complement the growing body of work focusing on the processes underlying interaction and meaning making in ELF situations.

References

Bauer, Laurie & Paul Nation. 1993. Word families. *International Journal of Lexicography* 6(4). 253–279.

Biber, Douglas & Susan Conrad. 2009. *Register, genre, and style*. Cambridge, UK & New York: Cambridge University Press.

Blommaert, Jan & Ben Rampton. 2011. Language and superdiversity. *Diversities* 13(2). 1–22. www.unesco.org/shs/diversities/vol13/issue2/art1 (accessed 30 August 2015).

Breiteneder, Angelika, Marie-Luise Pitzl, Stefan Majewski & Theresa Klimpfinger. 2006. VOICE recording – Methodological challenges in the compilation of a corpus of spoken ELF. *Nordic Journal of English Studies* 5(2). 161–188.

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.

Cogo, Alessia. 2010. Strategic use and Perceptions of English as a Lingua Franca. *Poznań Studies in Contemporary Linguistics* 46(3). 295–312.

Cogo, Alessia. 2012. ELF and super-diversity: A case study of ELF multilingual practices from a business context. *Journal of English as a Lingua Franca* 1(2). 287–313.

Cogo, Alessia & Martin Dewey. 2006. Efficiency in ELF communication: From pragmatic motives to lexico-grammatical innovation. *Nordic Journal of English Studies* 5(2). 59–93.

Corpus Description. 2013. The Vienna-Oxford International Corpus of English – Course Description. http://www.univie.ac.at/voice/page/corpus_description (accessed 30 August 2015).

ELFA 2008. The Corpus of English as a Lingua Franca in Academic Settings. Director: Anna Mauranen. http://www.helsinki.fi/elfa/elfacorpus (accessed 30 August 2015).

Firth, Allan. 2009. The lingua franca factor. *Intercultural Pragmatics* 6(2). 147–170.

Georgakopoulou, Alexandra & Dionysis Goutsos. 1997. *Discourse analysis: An introduction*. Edinburgh: Edinburgh University Press.

Gilner, Leah. 2014a. An analysis of ELF speakers' lexical preferences. *Asian English Studies* 16. 5–16.

Gilner, Leah. 2014b. Vocabulary preferences of English speakers in localized and globalized settings. Paper presented at the 4th Waseda ELF Workshop, Tokyo, Japan.

Gilner, Leah. 2015. Lexical distributions of high frequency words in spoken English as a lingua franca in academic settings. *Bunkyo Gakuin University Research Institute Journal* 14. 1–12.

Gilner, Leah & Franc Morales. 2011. The ICE-CORE word list: The lexical foundation of 7 varieties of English. *Asian Englishes* 14(1). 4–21.

Gilner, Leah, Franc Morales & Kayono Shiobara. 2011. The creation of a corpus of 26 international varieties of English. *Bunkyo Gakuin University Research Institute Joint Research Bulletin* 12. 35–43.

Gilner, Leah, Franc Morales & Kayono Shiobara. 2012. High frequency words in world Englishes. *Bunkyo Gakuin University Research Institute Joint Research Bulletin* 13. 99–111.

Gómez González, María de los Ángeles. 2013. A reappraisal of lexical cohesion in conversational discourse. *Applied Linguistics* 34(2). 128–150.

Hatch, Evelyn. 1992. *Discourse and language education*. Cambridge, UK: Cambridge University Press.

Jenkins, Jennifer, Alessia Cogo & Martin Dewey. 2011. Review of developments in research into English as a lingua franca. *Language Teaching* 44(3). 281–315.

Mauranen, Anna. 2006. A rich domain of ELF – The ELFA Corpus of Academic Discourse. *Nordic Journal of English Studies* 5(2). 145–59.

Mauranen, Anna. 2012. *Exploring ELF: Academic English shaped by non-native speakers*. Cambridge & New York: Cambridge University Press.

McEnery, Tony & Richard Xiao. 2005. Help or help to: What do corpora have to say? *English Studies* 86(2). 161–187.

Nation, Paul. 2001. *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, Paul. 2006. How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review* 63(1). 59–81.

Nelson, Gerald. 2011. The International Corpus of English. http://ice-corpora.net/ice/ (accessed 30 August 2015).

Palmer, Harold. 1931. *Second Interim Report on Vocabulary Selection submitted to the Eighth Annual Conference of English Teachers under the auspices of the Institute for Research in English Teaching*. Tokyo: IRET.

Pitzl, Marie Luise, Angelika Breiteneder & Theresa Klimpfinger. 2008. A world of words: Processes of lexical innovation in VOICE. *Views* 17(2). 21–46.

Ranta, Elina. 2006. The "attractive" progressive – Why use the -ing form in English as a lingua franca? *Nordic Journal of English Studies* 5(2). 95–116. http://hdl.handle.net/2077/3150 (accessed 30 August 2015).

Sinclair, John. 2004. Corpus and text – Basic principles. In M. Wynne (ed.), *Developing linguistic corpora: A guide to good practice*, 1–16. Oxford: Oxbow Books. http://ota.ahds.ac.uk/documents/creating/dlc/chapter1.htm (accessed 30 August 2015).

Seidlhofer, Barbara. 2009. Accommodation and the idiom principle in English as a Lingua Franca. *Intercultural Pragmatics* 6(2). 195–215.

Seidlhofer, Barbara. 2011. *Understanding English as a lingua franca*. Oxford: Oxford University Press.

Widdowson, Henry. 2012. ELF and the inconvenience of established concepts. *Journal of English as a Lingua Franca* 1(1). 5–26.

Widdowson, Henry. 2015. ELF and the pragmatics of language variation. *Journal of English as a Lingua Franca* 4(2). 359–372.

Bionote

Leah Gilner

Leah Gilner is Associate Professor in the Faculty of Foreign Studies at Bunkyo Gakuin University (Tokyo, Japan). Her research interests encompass word knowledge acquisition, applied phonetics and phonology, and language teaching methodologies. She serves as an editor for the *Asian Englishes* journal and her research has appeared in journals related to linguistics, language pedagogy, and sociolinguistics.