UNIVERSITY OF HELSINKI

https://helda.helsinki.fi

Usage-based recycling or creative exploitation of the shared code? : The case of phraseological patterning

Vetchinnikova, Svetlana

2015-09
De Gruyter Mouton
http://hdl.handle.net/10138/225288
Vetchinnikova, S 2015, 'Usage-based recycling or creative exploitation of the shared code? The case of phraseological patterning', Journal of English as a Lingua Franca, vol. 4, no. 2, pp. 223-252. https://doi.org/10.1515/jelf-2015-0019

Downloaded from Helda, University of Helsinki institutional repository. https://helda.helsinki.fi
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.
Please cite the original version.

Svetlana Vetchinnikova*

Usage-based recycling or creative exploitation of the shared code? The case of phraseological patterning

Повторение старого или творческий подход к общему коду? Что происходит при нестандартном употреблении устойчивого сочетания

DOI 10.1515/jelf-2015-0019

Abstract: This paper sets as its aim to show that ELF research is not uniform in its fundamental theoretical assumptions. While one wing of ELF scholars follows a classical Saussurean tradition with its notions of *langue* and *parole*, the other is distinctly usage-based. This is not presented as a problem but rather as a matter to be taken into consideration because the views are not easily reconcilable, if at all. Naturally, this divergence of view is clearly visible in data analysis and leads to different interpretations of the underlying processes. I will illustrate my point by the case of phraseological patterning. It is well-known that multiword units occurring in ELF use are not quite the same as in standard varieties. In SLA the discrepancies are traditionally labelled as errors, a stance ELF researchers do not generally agree with since the conflicting usages do not seem to lose their communicative successfulness. The question at which ELF researchers look differently is whether not quite nativelike phraseological patterning in ELF is a result of analytic or holistic – although fuzzy – processing, in other words, creation or approximation. The answer to this question has profound implications for our understanding of second language processing.

Keywords: usage-based models, classical Saussurean tradition, virtual language, phraseology in ELF, idiom and open-choice principles, Sinclair's model of a unit of meaning, approximation

Аннотация: Данная работа ставит своей целью показать, что исследования в области английского как лингва франка (АЛФ) не едины по своим теоретическим основаниям. Одна часть исследователей продолжает традицию классического соссюровского разделения языка и речи, а другая

***Corresponding author: Svetlana Vetchinnikova,** University of Helsinki, Helsinki, Finland, E-mail: svetlana.vetchinnikova@helsinki.fi

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM

отдает предпочтение современным моделям языка, основанным на его употреблении и отрицающим двуединую сущность языка, также, как и основополагающую роль системы языка, существующей в абстрактной форме. Это не проблема, но соображение, которое следует учитывать, т.к. данные позиции трудно, если не невозможно примирить. Естественно, что несовпадения во взглядах ясно прослеживаются в анализе данных и приводят к различным интерпретациям происходящих процессов. Наша точка зрения будет раскрыта на примере фразеологии, которая в данной работе трактуется широко: под фразеологической единицей понимается любая языковая единица, состоящая из нескольких слов и обладающая определенной степенью устойчивости. Широко известно, что фразеологические обороты в АЛФ отличаются от их употребления в стандартных вариантах языка. В научной области, изучающей усвоение английского как второго языка, такие отклонения принято называть ошибками, тогда как исследователи АЛФ с этим не вполне согласны, поскольку подобные отступления от узуального употребления не препятствуют достижению целей коммуникации. Мнения исследователей АЛФ расходятся в том, что касается не совсем похожего на носителей языка употребления фразеологических единиц: являются ли они результатом творческого переосмысления и вторичной сборки из частей или всего лишь результатом аппроксимации, или приблизительного воспроизведения, объясняющегося особенностями памяти и влиянием частотности употребления. Ответ на этот вопрос влечет за собой далеко идущие последствия для понимания процессов порождения речи на втором языке.

Ключевые слова: модели языка, основанные на частотности употребления, соссюровская традиция, виртуальная форма существования языка, фразеология в английском как лингва франка (АЛФ), принцип идиоматичности и принцип свободного построения фразы, модель единицы значения Синклера, аппроксимация

1 Introduction

Is second language (L2) processing fundamentally different from first language (L1)? It is evident that L2 use[1] is different from native language

1 In this article I often talk more broadly about L2 users since I am interested in cognitive processing and its manifestation in language use. Since most ELF speakers are cognitively L2 users, all the arguments equally apply to ELF use as well (see Mauranen 2011 for a

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM

use.[2] In SLA research, the differences are traditionally seen as deficiencies of non-native language performance which need to be remedied to assist non-native speakers (NNSs) in approaching native speaker (NS) performance as closely as possible. But how do we conceptualise language use which diverges from the standard but is still communicatively successful? It cannot be seen as an imperfect learner language going through a developmental stage. The need to account for this type of language use inspires ELF research. Thus, many ELF scholars look at language use in ELF settings in its own right, from the perspective of communicative successfulness, legitimising (some of the) departures from the standard. However, I argue that ELF studies put forward two distinct theoretical explanations of how these differences come about which are mutually exclusive. I will try to show that their roots go very deep into the fundamental conception of the nature of language and in fact belong to different paradigms of linguistic scholarship which are not limited or specific to ELF studies in particular.

So, one explanation of the differences between NS and ELF use is formulated in a classical Saussurean tradition with its distinction between linguistic code, an underlying abstract set of rules, and language usage, while the other follows a usage-based approach to language acquisition, use, and change. Classical and usage-based approaches to language, being fundamentally different, are in direct opposition to each other and often engage in open debate (see, e.g., Newmeyer 2003 and responses to it in *Language*). There are numerous issues in linguistics on which these two paradigms of thinking about language do not agree, from first language acquisition (see Tomasello 2003) to grammaticalisation (see, e.g., van Gelderen 2011). One of such issues is the role of corpora in linguistic research which has been especially controversial (see, e.g., Aarts 2001; Widdowson 2000; Stubbs 2001; McEnery et al. 2006: Unit B2; McEnery and Hardie 2013). This is of course not surprising since the classical position sees language use, which is what is compiled in corpora, as secondary and focuses on the properties of the abstract underlying system instead, while for usage-based scholars language use is the only form in which language exists and so grammar is a description of language in use.

discussion of the differences and the commonalities between L2 users, including ELF speakers, and L2 learners).

2 NS versus NNS might not be a particularly useful dichotomy (not least because neither of the two groups are homogeneous in their language skills), especially when it comes to English used as a lingua franca (see, e.g., Jenkins 2000; Hynninen 2014), but it is the one most widely drawn. Thus, it seems important to deal with it and try to understand the reasons behind the apparent differences.

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM

In this paper I will discuss the two co-existing explanations of the differences between NS and ELF use with regard to phraseological patterning, which is in many ways key to understanding language processes as it is now widely accepted to be at the core of language (see, e.g., papers in Granger and Meunier 2008). According to the first explanation, the not quite nativelike multi-word units (MWUs) produced by L2 users in ELF settings are novel word combinations which demonstrate creativity of ELF speakers in their exploitation of latent possibilities inherent in the virtual language (in contrast to conformist ENL actualisations of the already encoded possibilities).[3] According to the second, such MWUs are not novel despite formal variation: they exhibit patterning common to the kind of English (in terms of register and/or variety) these ELF speakers were exposed to. I argue that in each particular case these two explanations are mutually exclusive: either the use of an MWU is novel and therefore can be described as creative or it is not. The notion of creativity is taken here in its most basic sense of implying a production of something new as an explicit and independent choice contrasted with an implicit reuse of linguistic material conditioned by previous experience with the language.[4] The distinction between creative and "non-creative" uses can then be formulated using Sinclair's idiom and open-choice principles (Sinclair 1987, Sinclair 1991), which has indeed been done in previous studies (Seidlhofer 2009; Pitzl 2009: 313). From this perspective, an MWU can either be produced on the open-choice principle by selecting each word independently and combining them online using the rules of grammar, in which case it is at least in some way novel and therefore can be described as creative, or it can be produced on the idiom principle by reliance on memorised instances of the same pattern experienced before, in which case it is not novel. When the phraseological competence of L2 users is looked at on the whole, it is the extent to which they construct novel combinations of words versus reuse the patterns experienced, i.e., operate on the open-choice versus on the idiom principle, which is in question. In the case of native language use, it has been repeatedly shown that while both operating principles are in use, idiom principle is the default (e.g., Erman and Warren 2000). What about ELF speakers then? Do they differ from native speakers in that they primarily rely on the open-choice principle?

3 I use *multi-word unit* (MWU) as a generic term to refer to all kinds of phraseological patterning.

4 The distinction between explicit and implicit here is the same as between explicit and implicit learning (also called declarative and procedural) or explicit and implicit memory systems. So explicit here is synonymous to conscious, rather than deliberate or intentional which may or may not constitute the defining characteristic of creativity (see the discussion in Pitzl 2012: 30–33) but is not an issue in this paper.

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM

In what follows, I will first describe the two directions of thinking in ELF research and try to explain why they can be viewed as belonging to different paradigms of linguistic scholarship (Section 2). In Section 3, I will move on to specify the positions of ELF scholars within these two directions of thinking with regard to phraseological patterning in ELF. It is assumed that there is a link between the main tenets of the paradigm, different in each case, ELF researchers subscribe to or naturally inherit and their explanation of the reasons behind divergent phraseological patterning in ELF. In the empirical part of the paper, I will adopt a usage-based perspective and try to show in which way phraseological patterning in L2 is not novel but based on personal experience with the language. In Section 4, I will present some quantitative findings obtained by comparing L2 writing to the language these L2 writers were exposed to. In Section 5, I will zoom in on some of the non-matching patterns identified in the comparison to see whether these patterns can be viewed as novel word combinations or whether they exhibit reliance on experienced language despite formal variation. The results and implications are summarised in the concluding section.

2 Classical Saussurean versus usage-based view on language

As I have suggested above, there are two distinct theoretical frameworks in which current ELF research develops and which provide different explanations of why ELF use departs from NS use. According to the first view, most notably held by Widdowson and Seidlhofer, actual language, or language as we encounter it in use, is an instantiation or realisation of the virtual language. More precisely, it is only one of the possible instantiations of the virtual language and therefore by definition does not realise all the potential which is contained in it. Thus, any of the "core" varieties of English, for example, American English or British English, are the already encoded instantiations of the possible. The codified linguistic items, whether single or multi-word units, just happen to be the ones realised out of all the possibilities virtual English permits. It is for this reason that ELF variants can be seen just as legitimate: they rely on the same virtual English and instantiate the possibilities contained within it, only not (always) the ones already instantiated in "core" varieties. In other words, instead of reproducing the forms established in "core" varieties, ELF users draw on the virtual language to create new forms, which, with time, can very well replace the established ones. Hence, there is an emphasis on creation (of

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM

novel forms) and creativity, manifest in such concepts as idiomatising and languaging, which characterises this account of ELF.

Let us now focus on the concept of a virtual language. The concept is suggested in Widdowson (2003, see also Widdowson 1997) where he compares the way writers of literary works "exploit the resources of a language" to nonconformities of learner language, arguing that in both cases a novel combination produced, "not allowable by the conventional code," is "nevertheless a latent possibility which is *virtual* in the language though not actually encoded" (Widdowson 2003: 48, emphasis in the original). In particular, Widdowson writes:

> We cannot just dismiss all examples of nonconformist usage as aberrations or oddities, no matter how appropriate to purpose they may be. It would be preferable, I suggest, to consider them as evidence of the existence of what I shall call the *virtual* language, that resource for making meaning immanent in the language which simply has not hitherto been encoded and so is not, so to speak, given official recognition. If we did not recognize this virtual reality, we would not be able to make any sense of these nonconformities at all. (Widdowson 2003: 48–49)

The concept of the virtual language is further elucidated in terms of constitutive and regulative rules, following Searle (1969, 1995). While constitutive rules define the game being its code, regulative rules "only characterise different ways of playing it" (Seidlhofer 2011: 113). Constitutive rules are categorical and invariant, they bear the status of actual rules; regulative rules are variable, they are merely conventions of playing the game. As such, virtual language is an underlying system of constitutive rules (see Seidlhofer 2011: 112). So, for example, what corpus linguists are able to observe in corpora is only the regulative rules, language users' performance rather than their competence (Widdowson 2012), and in that sense Seidlhofer and Widdowson see corpus linguistics as "reductionist" (Seidlhofer 2011: 113)[5] or "only a partial account of the real language" (Widdowson 2000: 7).[6] But at the same time, "it is

5 "[...] it has become customary to criticize sentence grammars of the traditional kind as reductionist, arid abstractions that do not account for language in performance, the actual language. But then they were never designed to do so. And we need to note that descriptions of actual language based on corpus data are also reductionist in that they fail to account for the potential inherent in the virtual language" (Seidlhofer 2011: 113).

6 "We get third person facts of what people do, but not the facts of what people know, nor what they think they do [...] we can say that corpus analysis deals with the textually attested, but not with the encoded possible, nor the contextually appropriate [...]. What it [corpus analysis] can do is reveal the properties of text [...]. But it is necessarily only a partial account of the real language. For there are certain aspects of linguistic reality that it cannot reveal at all. In this respect, the linguistics of the attested is just as partial as the linguistics of the possible" (Widdowson: 2000: 6–7).

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM

entirely possible that certain patterns of regulative behaviour through repetition, become so established over time that they become ratified as constitutive rules of the game," Seidlhofer (2011: 114) continues. This is how change is made possible in this account.

The distinction between constitutive and regulative rules serves well in interpreting non-conformities observable in ELF data and legitimising them. Seeing only native-speaker usage "as the real language" would mean "to elevate a particular set of regulative conventions to the status of constitutive rules" and to assume "not only that the native-speaker way of playing the English game is the best way, whatever the conditions or circumstances, but that it is really the *only* way" (Seidlhofer 2011: 114, emphasis in the original). Therefore, we need to shift the focus from the regulative conventions to the constitutive rules of the virtual English on which both native-speaker English and ELF are based. Then it becomes clear that ELF usage is just as legitimate as ENL.

There is a different conceptualisation of ELF which takes a usage-based perspective on language widely adopted in linguistic research at large (see, e.g., Bybee 2008, Bybee 2010; Ellis 2012; Tomasello 2003). I will build my account on Mauranen (2012). In short, this position is very simple: language emerges from usage rather than exists independently of it, therefore English used as a lingua franca, as language use characterised by distinct conditions (e.g., presence of other languages, global spread), shall have emergent properties different from, e.g., native monolingual varieties. It assumes that the resource language users resort to is not the virtual language or a shared "linguistic code" (Kecskes 2007) but their personal accumulated experiences of this language. Each individual's accumulated experience has a cognitive representation which is constantly changing as new experience builds up, a position which is in agreement with exemplar models of language representation (see, e.g., Bybee 2006). At the communal level, language arises in interaction between different personal experiences of it (see Beckner et al. 2009 and Ellis 2011 for the distinction between the communal and the individual levels of language as a feature of a complex adaptive system). At this communal level we can further distinguish between a microsocial and a macrosocial perspective (Mauranen 2012: Ch. 2): language innovations originating in natural human interaction feed into the language seen from a macrosocial perspective through diffusion or propagation (see Mauranen 2012: 26).

A major emphasis in a usage-based approach is placed on frequency and probability: the more frequently an item is encountered, the stronger its cognitive (exemplar) representation and the more probable is its further use. The more frequent is the item overall at the communal level, the greater is its chance to be preserved in language unchanged (see Mauranen 2012: 33–34). For this reason,

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM

corpus linguistic methods of investigation are put high on the agenda since by counting frequency and probability they allow language regularities to be observed. On the whole, observation of language in use is essential in this approach because it is considered to be the only form in which language exists. This view is compatible with the theories of language emergence (e.g., Hopper 1987; Bybee and Hopper 2001a; Ellis and Larsen-Freeman 2006) and language as a complex adaptive system (e.g., Beckner et al. 2009; Ellis 2011; Larsen-Freeman 2013).

If we now go back to the idea of the virtual language with its insistence on a clear distinction between constitutive and regulative rules and a sharp dividing line between knowledge of language and use of language, it seems reasonable to argue that it continues the classical Saussurean tradition with its notions of *langue* and *parole* as well as Chomskyan distinction between competence and performance, "Internalised language" or "I-language" and "Externalised language" or "E-language" (Chomsky 1986). Viewed this way, the difference in the positions of ELF scholars is part of a larger debate, as suggested in the introduction, which has a long history (see, e.g., Newmeyer's [2003] article "Grammar is grammar and usage is usage" in *Language* and the heated discussion it generated in Laury and Ono 2005; Meyer and Tao 2005; Bybee and Hopper 2001b). Is there an abstract linguistic system/competence and does it underlie all usage/performance? Or is competence merely an abstraction from all instances of usage/performance and therefore does not really exist as such, being a construct which we need in order to talk about one's future language use predicted by his/her previous language use? In other words, what comes first, system/competence or usage/performance? A kind of chicken and egg problem. This paper does not pursue the ambition of solving this problem and settling the debate. The only modest goal it has is to make explicit the difference in the views of ELF scholars, which plays a crucial role in how they conceptualise ELF, how they collect, analyse, and interpret data.

3 The case of phraseological patterning

With regard to the two approaches, the divide between them is perhaps most apparent in their respective analyses of multi-word patterning.

Not quite native-like phraseological patterning of second language use has attracted the attention of scholars for over three decades, roughly from Pawley and Syder's (1983) widely known article on "nativelike selection and nativelike fluency." The common conclusion researchers come to is that phraseological patterning or, simply, multi-word units (MWUs) of all types constitute a major

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM

difficulty for second language users even at advanced levels of proficiency (Granger 1998; Howarth 1998; papers in Schmitt 2004, Meunier and Granger 2008). In contrast to NSs, they simply do not get the phrases right, in SLA researchers' view, as can be seen from the examples available in abundance in SLA literature, some of which are presented below (example [1]):

(1) a. (Yorio 1989: 62–63)
take advantages of (take advantage of); a friend of her (a friend of hers); make a great job (did a great job); on the meantime (in the meantime); with my own experience (in my own experience); put more attention to (pay)
b. (Howarth 1998)
draw a conclusive comment (draw a conclusion); do attempts (make); get contact with (make contact with/get in contact with); suffer difficulty (suffer pain/experience difficulty)
c. (Nesselhauf 2005)
reach an aim (achieve); make (a) career (have); open up jobs (create); bring about solutions (find); stick to a prejudice (cling to); fail at their task (in)

Therefore, it is not surprising that ELF researchers actively engaged in this line of research, too. In fact, the types of infelicities noticed in ELF use are very similar to learner language: after all, cognitively both types of use represent second language use (see Mauranen 2011). So just as L2 learners, ELF speakers introduce structural/grammatical (example [2]) and semantic/lexical changes (example [3]) into standard phraseology.

(2) a. (Mauranen 2009, Mauranen 2011)
in/on my point of view (from); on this stage (at); to put the end on it (to); take closer look to the world (at); on the end (in/at)
b. (Carey 2013)
in the same moment (at)
c. (Seidlhofer 2009)
on the base of my experience (in)
d. (Pitzl 2012)
on the long run (in); carved in stones (stone), pieces by pieces (piece by piece)

(3) a. (Mauranen 2011, Mauranen 2012)
*the hen or the egg (chicken); a few / a couple of / some words about (*versus *a few words about* only in MICASE*)*

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM

b. (Carey 2013)
on the other side (hand); so to say (speak)
c. (Seidlhofer 2009)
in my observation (view); to my head (mind)
d. (Pitzl 2012)
draw the limits (line); preserve their face (save); keep in the head (keep/bear in mind); turn a blank eye (blind)

ELF researchers do not problematise these divergent uses because in all cases it is found that deviations from standard phraseology do not result in communicative turbulence, and therefore variant forms can be considered communicatively successful. It is also argued that on the whole it might be a sensible strategy for an ELF speaker to avoid using opaque NS idioms since this is likely to lead to "unilateral idiomaticity," a situation where interlocutor(s) do not recognise the idiom which makes the whole utterance confusing (Seidlhofer 2011). But this is where the consensus ends. The disagreement exists with respect to the explanation of how phraseological units with non-standard features come about in the first place.

SLA researchers have long been arguing that the reason why L2s make mistakes in phraseological patterning lies in the processing differences between L1 and L2: while L1s just retrieve the patterns holistically from memory producing them on the idiom principle, L2s build them analytically from parts, i.e., on the open-choice principle, which is how errors creep in and why the end result is different from the target standard phraseology (see, e.g., Wray 2002; Granger 1998). Seidlhofer, Widdowson, Pitzl, and Kecskes (Seidlhofer and Widdowson 2007; Seidlhofer 2009, Seidlhofer 2011; Pitzl 2009; Kecskes 2007) seem to agree with this view. For example Seidlhofer and Widdowson (2007: 365) hold that ELF users "construct what they have to say more atomistically, in a bottom-up fashion." Developing the argument, Seidlhofer (2009: 205) writes that ELF users experience "a primary dependency on the open-choice principle" due to limited knowledge of phraseological patterns:

> They [users of ELF] may have some of the most commonly occurring fixed phrases available to them (*of course, on time, by the way*) but often not those which native speakers have intuitive access to [...]. This being so, ELF users will tend to construct what they have to say more analytically, in a bottom-up fashion, drawing on what is semantically encoded in the grammar and lexis of the language – in other words, by recourse to what Sinclair calls the "open-choice" principle. [...] One needs to recognize that the mode of communication among ELF users may be quite radically different from that which characterizes the linguistic behaviour of ENL speakers [...]. (Seidlhofer 2009: 202)

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM

In a more or less similar vein, Kecskes (2007) argues that due to the lack of shared background, ELF users have to rely on the "linguistic code" and use it "as directly as possible" which for him means using words with their literal meaning: "avoid[ing] genuine formulaic language and [giving] preference for semantically transparent expressions" (Kecskes 2007: 207, 212).

Pitzl (e.g., Pitzl 2009, Pitzl 2012) examines instances of creative idioms, that is, idioms which exhibit certain variation: lexical, morphosyntactic, syntactic, or a combination of these, in spoken ELF data using the VOICE corpus. She provides very interesting examples, for instance: *we should not wake up any dogs* (instead of *Let sleeping dogs lie*), *put my hands into the fire for it* (cf. *de hand voor iemand in het vuur steken* in Dutch, the speaker's L1), *smooth the process* (*path/way*), *remember from the head* (*off the top of your head)* (see also examples [2] and [3]). Even though all these instances fall short of the NS standard, so to speak, they still seem to be entirely communicatively successful. Pitzl (2012: 48) explains observable variation by "a natural pull towards non-conformity and self-expression" which makes ELF speakers approach NS idioms creatively and re-activate the sleeping metaphorical potential in them. She also takes this formal variation to be clear evidence of their compositionality and ensuing semantic transparency. This suggests, she concludes concurring with Seidlhofer (2009), that "ELF speakers seem to construct what they say more atomistically." Analysing the expression *how to draw the limits*, which was produced instead of the conventional *to draw the line*, she suggests the following interpretation:

> Rather than being stored, processed and uttered as a single lexical unit, i.e. via the idiom principle, the expression *how to draw the limit* seems to be created via the open-choice principle. The semantic properties of the individual words are foregrounded and exploited, which is evident in the word *limit* being substituted for *line*, both words denoting similar semantic concepts. Such a formal variation of an otherwise conventional idiomatic expression reactivates metaphorical resonance. (Pitzl 2009: 313)

There are two issues I would like to point out in Seidlhofer's, Kecskes's, and Pitzl's position. The first is a characteristic focus on fixed phraseology only. Phraseology is narrowly interpreted as idioms or phrases with almost entirely non-compositional meaning. For example, in his study, Kecskes (2007) counted the occurrence of formulaic expressions in a 13,726 word database of ELF conversations which reached only 7.6% of the total words. This is reasonably estimated as very little, especially in comparison to seemingly similar counts in NS language use (Erman and Warren [2000] report about 50%, Altenberg [1998] even 80%, while Dąbrowska [2004] predicts that in reality the percentages are likely to be higher). Therefore Kecskes (2007) concludes that "[n]on-native

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM

speakers appear to rely on prefabricated expressions in their lingua franca language production to a much smaller extent than native speakers" (Kecskes 2007: 199) and that "multiword chunks might not help L2 processing in the same way they help L1 processing" (Kecskes 2007: 202). However, it needs to be taken into account that in his analysis Kecskes (2007: 194) "concentrated only on fixed expressions that are usually motivated and allow relatively few structural changes," as he explicitly states himself. Arguably, such expressions constitute only a minor part of the "phraseological tendency" of language (see, e.g., Cheng et al. 2009), and therefore cannot be used to make claims about the extent to which NNSs rely on prefabricated expressions overall or the extent to which multi-word chunks might be helping or not helping L2 processing (this is not what Erman and Warren [2000], Altenberg [1998], Dąbrowska [2004] counted either). How about patterns like *a large/small/growing/limited/considerable number of* or *it is generally/widely/commonly/usually assumed that* (BNC) or *It is* adj + *to*-inf as in *It is reasonable to argue* / *It is important not to underestimate*, which exhibit systematicity despite wide variability? In this paper I pursue a wider approach to phraseological patterning and thus adopt the model of a unit of meaning developed by Sinclair since it allows for the variability excluded from other models (see Section 5.1).

The second issue is the interpretation of variation observable in ELF use of multi-word units. It is assumed to be a sign of open-choice processing, the idea being that if they were retrieved holistically, alterations would not occur. However there is a different interpretation of the variation. Mauranen (2005, 2009, 2011) analyses very similar examples from the ELFA corpus, for instance: *in/on my point of view* (versus *from), on this stage* (versus *at), to put the end on it* (versus *put an end to it), the hen or the egg* (versus *chicken)* (see also examples [2] and [3]) and draws attention to the fact that these altered variants are not only communicatively successful but also occur across different speakers which rules out the possibility of idiosyncratic creativity. "What is thus worth noticing about these units," she writes, "is that ELF speakers [...] get them approximately right," rather than slightly wrong (Mauranen 2012: 144).

And indeed if we take a usage-based perspective with its idea of accumulating exemplar representation (see, e.g., Bybee 2006, Bybee 2010), low frequency of exposure and consequent weak entrenchment may very well lead to semantic and structural approximations in the formal realisation of a lexical unit which we often observe in ELF use. Memory for meaning is stronger than memory for surface structure (e.g., Gurevich et al. 2010), therefore, to take Pitzl's example *how to draw the limits*, we could hypothesise that *limits* was substituted for *line* because rather than being constructed compositionally, the unit was

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM

remembered holistically but formally incompletely: instead of a specific word *line*, the memory was of a semantic feature of some kind of "borderline" or "boundary." Language use does not allow speaking in semantic features: an item representing the required semantic feature must be selected and produced. In this particular case, it happened to be *limits*, but it could just as well be *border* or *boundary*, though it is probably not by chance that *limit* and *line* start with the same letter – do we not often remember the starting letter of a word we otherwise forget? So the substitution, in this view, occurs as a result of *approximation*, or inexact memorisation, presumably typical of fuzzy processing (see Mauranen 2012: 42–44), rather than creative construction.[7]

So, is it the idiom or open-choice principle which predominantly underlies ELF and, in more broad terms, L2 use? Do L2 users to a larger extent display usage-based recycling of the patterns previously experienced or creative exploitation of a shared language resource in their language production?

4 Idiom principle in L2 acquisition and use: some quantitative findings

This section presents some findings of a larger study (Vetchinnikova 2014) which are relevant for the discussion of the questions posed in this paper in that they can shed light on the extent to which L2 users recycle experienced language patterns in their own production. In this larger study, I examined the availability of the idiom principle to L2 users in language acquisition, processing, and use. As part of the study, I collected five students' drafts of Master's theses written over a period of about one year and compiled them into an individual corpus of written production for each student (henceforth, Production Corpus). To probe acquisition of MWUs from exposure and create a viable point of comparison for the texts students produced, I collected academic publications each student referred to in her thesis into language exposure or priming corpora (henceforth, Exposure Corpus). Table 1 shows the sizes of the resulting corpora.

7 While I would insist that lexical and structural substitutions are likely to be introduced into L2 phraseology on the idiom principle, I admit that such substitutions may occasion a switch to processing on the open-choice principle and therefore re-metaphorisation on the part of the hearers. In other words, inexact reproduction of the form of a conventional idiom might be re-metaphorising it for those hearers/readers who are used to its conventional form and will therefore process the substituted part as unusual co-text.

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM

Table 1: The size of each student's Production Corpus and Exposure Corpus.

Student	Kaisa[a]	Hertta	Maisa	Linda	Nora
Production Corpus	21,887	39,449	21,501	37,641	14,843
Exposure Corpus	64,809	123,662	85,234	133,778	86,500
Exposure Corpus/Production Corpus[b]	3.0	3.1	4.0	3.6	5.8

[a]Students' names are pseudonyms.
[b]The size of Exposure Corpus divided by the size of Production Corpus.

A comparison between these two types of corpora for each user is able to clarify to what extent L2 use is a result of usage-based priming and to what extent it has to rely on novel combinations of words. In other words, do L2 users acquire MWUs holistically from exposure or do they create them ad hoc on the open-choice principle? What is the distance between the patterning of a priming language and the patterning of L2 production? If L2 users are forced to rely on the "linguistic code" and build on the semantics of individual words in their constructions, the overlap between the multi-word patterning of the language they are exposed to (Exposure Corpus) and the patterning of their own language (Production Corpus) will not be significant and will occur by chance only. If, instead, the comparison shows that there is substantial overlap and continuity between the two corpora, this will suggest that L2 learning is usage-based and happens through exposure.

To find out how much the MWU patterning of Production Corpus had in common with that of Exposure Corpus, I used the following procedures. First, I ran both types of corpora through a phraseological engine ConcGram (Greaves 2009). ConcGram is able to retrieve all two-word co-occurrences (concgrams) irrespective of positional or constituency variation (see Cheng et al. 2006). So for example, it can register the co-occurrence of PLAY and ROLE, not only in *play a role* or *play an important role* but also in *role play* and *a leading role to play* (Cheng et al. 2009).[8] In this study, I set the internal window span at 4 and minimum frequency at 2, that is, asked the program to extract all two-word combinations which occur more than once with up to four intervening words in between. Not surprisingly, a lot of meaningless two-word combinations,[9] like *of the*, seeped through. To filter the resulting lists, I used statistical significance measures and set the cut-off points at *t*-score = 2 and MI value = 3 (see McEnery et al. 2006:

8 In this paper I use SMALL CAPS for concgrams and lemmas and *italics* for word forms.
9 "Meaningless" here is used in the sense of concgrams not having an autonomous meaning of their own rather than being nonsensical.

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM

56–57; Hunston 2002: 71–72). When I had lists of statistically significant concgrams for each of the student's Production Corpus ready, I asked the program to compare them to the corresponding Exposure Corpora. The program returned the lists with information on how often each of the concgrams from a Production Corpus list occurred in the corresponding Exposure Corpus, including zero values. Thus, it was possible to calculate the percentage of overlapping concgrams, i.e., Production Corpus concgrams which also occurred in Exposure Corpus at least once. The results can be seen in Table 2.

Since the resulting percentages of overlap, while being clearly significant, are not self-evidently high or low, I then compared a concgram list of each student's Production Corpus with a different student's Exposure Corpus, i.e., Exposure Corpus representing the language of a different field of study from the one the student came from. Table 3 displays the result.

Table 2: The percentage of matching patterns between Production Corpus and Exposure Corpus.

Student	Significant concgrams ($t \geq 2$ & MI ≥ 3), Production Corpus	Number of Production Corpus significant concgrams ($t \geq 2$ & MI ≥ 3) which also appear in Exposure Corpus	Percentage of Production Corpus significant concgrams ($t \geq 2$ & MI ≥ 3) which also appear in Exposure Corpus
Hertta	464	271	58
Kaisa	347	194	56
Linda	524	331	63
Maisa	260	196	75
Nora	108	74	69
Average			64

Table 3: The percentage of matching patterns between non-corresponding Production Corpus and Exposure Corpus.

Production Corpus	Exposure Corpus	Number of Production Corpus significant concgrams ($t \geq 2$ and MI ≥ 3) which also appear in Exposure Corpus	Percentage of Production Corpus significant concgrams ($t \geq 2$ and MI ≥ 3) which also appear in Exposure Corpus
Hertta	Maisa	92	20
Kaisa	Linda	70	20
Linda	Hertta	171	33
Maisa	Nora	74	28
Nora	Kaisa	33	31
Average			26

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM

As can be seen from Table 3, the overlap of concgrams between Production Corpus and a non-corresponding Exposure Corpus is much lower. This bears out the expectations, since while different fields of study share some patterns, they also develop field-specific ones. What is important for us here is that the patterning of students' writing matches the patterning of language in their own field better than in a different field, which suggests that the overlap between the corresponding Production Corpus and Exposure Corpus is significantly high. This suggests that L2 acquisition of MWUs is usage-based and happens on the idiom principle to a much greater extent than often claimed by other scholars in the field.

5 From exposure to use: a qualitative analysis

To examine the predictions of the usage-based approach further, I would like to now probe the proposed continuity between experienced language and produced language qualitatively. For this purpose, I will take several concgrams from the students' Production Corpus which did not match any pattern in their corresponding Exposure Corpus and analyse them using Sinclair's (1996) model of a unit of meaning. None of these examples have been previously analysed for the larger study mentioned above. In what follows, I will first describe the model and then proceed to the analysis itself.

5.1 Sinclair's model of a unit of meaning

In this paper I have chosen Sinclair's model of a unit of meaning (Sinclair 1996, Sinclair 2004) for the analysis of phraseological patterning in L2. This model allows for maximum variability inside the unit compared to other models I know of and is therefore particularly suitable for exploration of variation and change in the usage-based framework. It seems to be psycholinguistically real, too (Vetchinnikova 2014), or, in other words, there is a cognitive basis which can account for the components of the model.

A unit of meaning or a meaning-shift unit (MSU), as Sinclair calls it in his later works (Cheng et al. 2009), is essentially a more abstract form-meaning pairing than a word which is what seems to unite it with a notion of a construction developed in different branches of construction grammar (e.g., Goldberg 1995, Goldberg 2006; Croft 2001). It has two obligatory components: the core, a formal component, and the semantic prosody, a meaning-component. The core

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM

is what makes different configurations of the same unit of meaning formally recognisable as its variants. The semantic prosody is a functional meaning a unit consistently communicates despite formal variation. The major advantage of the model lies in its facility for optional components: a verbatim association – collocation – and two abstracted associations – semantic preference (a semantically abstracted association) and colligation (a structurally abstracted association). Psycholinguistically, this means that syntagmatic association can form not only between two specific words, like in *strong tea*, but also between a word and a semantic feature (or a semantic set of words), like in *to (a/an) some/larger/lesser/great/certain/considerable/* (etc.) *extent* where there is a semantic preference for a modifier specifying the "importance" or "size" of the *extent*, as well as a word and a grammatical feature (or a grammatical class of words), like in *when it comes to* + NP, where a NP has to follow but it is not semantically restricted in any way. In any case, there is an overall functional meaning which can be identified for every MSU despite different formal realisations of its components. For example, the unit *to a large extent* with its variants serves to evaluate the "degree to which something is or is believed to be the case" (ODE, *extent*).

Thus, a unit of meaning is a lexical item, usually consisting of more than one word, produced holistically, "by remembering" rather than "by putting together" (Bolinger 1976: 2), that is, on the idiom principle. In fact, a sequence of words becomes a unit when it is produced on the idiom principle because it is at this point that the hitherto separate words delexicalise (lose their core meanings) and acquire a new holistic meaning, semantic prosody – a process which constitutes a meaning-shift referred to in the new term, MSU. The more fixed the unit becomes with frequent use, the more its component words delexicalise and the more non-compositional the overall meaning of the unit is. So meaning-shift, while at first occasioned by the switch from the open-choice principle to the idiom principle, is thereafter a gradual process producing a continuum of delexicalisation. Can the process be reversed? Yes, it can: a unit of meaning can move back along the continuum becoming less fixed and yet retaining its status of a unit. This might happen due to, e.g., receding frequency of occurrence or other factors. In this case, collocation can become a semantic preference or a colligation. That is, verbatim association can be approximated, to use Mauranen's concept of *approximation*, first suggested in Mauranen (2005) and further developed in subsequent publications. In contrast to this process of mere loosening of the pattern without it breaking up into separate words, a unit of meaning can also be relexicalised, that is, reinterpreted compositionally on the open-choice principle, a process which seems to be similar to Pitzl's proposal of *re-metaphorization* in idioms (Pitzl 2009, Pitzl 2012). Forced relexicalisation is a technique used in word play (see Partington 2006) where the initial

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM

interpretation on the idiom principle is clashed against a compositional interpretation for ironical or comical effect.[10] It follows then that where there is no meaning-shift from the original semantic prosody of a unit of meaning, there has been no switch from the idiom principle to the open-choice principle: the structure of a unit might have loosened formally but it has not dissolved into component words – it is still the same unit.

6 Exposure Corpus versus Production Corpus

In what follows, I will analyse several MWUs, operationalised as concgrams, from the L2 users' Production Corpora which did not match the patterning of the corresponding Exposure Corpora, as revealed by the automatic comparison of concgrams in Production Corpus and Exposure Corpus explained in Section 4. I focus on the non-matching patterns since matching patterns already support the usage-based prediction of language acquisition from exposure and recycling of experienced patterns. It is more interesting to see what happens when the patterns do not match and how these divergent uses can be explained. I will analyse the changes L2 users introduced into the divergent MWUs by applying the model of a unit of meaning. In cases of inexact reproduction of an attested unit, my goal will be to establish whether meaning-shift has occurred, which would give grounds for the open-choice explanation, or whether there has been no such meaning-shift and all the alterations to the original unit could have occurred due to approximation.

As my first example, I take the concgram AMOUNT/DEATHS which occurs in Maisa's Production Corpus in the pattern *amount of (AIDS) deaths* five times (example [4]) but does not appear in her Exposure Corpus. The reason is that Exposure Corpus writers use *number of (AIDS) deaths* instead (example [5]).

(4) (Maisa, Production Corpus)

1 if there were no antiretroviral treatment. The **amount** of **death**s, on the contrary, is on the rise,
2 of AIDS, as expected. In 2020, the projected **amount** of **death**s is around 31,000 in declining and
3 2020. The reason for this is most probably the **amount** of AIDS **death**s, which would increase if there
4 Figure 16. Total births, 2001–2020. The **amount** of total **death**s (figure 17) is growing because of
5 an increase of only 5%. Here, the considerable **amount** of AIDS **death**s can be seen. (Ibid., 11.) The

10 For example:
A: What happens if the parachute doesn't open?
B: That's known as "jumping to a conclusion". (Partington 2006: 119)

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM

(5) (Maisa, Exposure Corpus)

of deaths will grow and eventually approach the **number** of **death**s prior to treatment, though patients
with the continued rapid roll out of ARVs, the **number** of **death**s per year due to AIDS will likely
treatment population increases, the absolute **number** of **death**s will grow and eventually approach the
in the 12th month category, compared with the **number** of **death**s in the 0–11 month categories, and
in Namibia. The scale of HIV infections, the **number** of **death**s that will occur mainly in the main
living with HIV AIDS by age and sex, the **number** of AIDS **death**s, and the number of orphans as a
people living with HIV AIDS by age and sex, the **number** of AIDS **death**s, and the number of orphans as a
projected ART coverage. Then you can display the **number** of AIDS **death**s to see the effect of expanded
of death in 2005 indicated that the annual **number** of registered **death**s rose by 87% between 1999

Amount of is a quantifying expression that is used with non-count nouns but informally can also occur with plural count nouns (Carter and McCarthy 2006: 348). So strictly speaking, it is not entirely conventional to say *amount of deaths*, and *number of deaths* would be a preferred variant. In fact, Maisa uses *number* in this pattern, *number of deaths* (7 instances), as well as in a larger pattern *number of* + count noun, too. *Amount of* + count noun occurs in her texts 14 times. To compare, in Exposure Corpus, *number of* + count noun has 183 instances, *amount of* + count noun – 2 instances.

The substitution of *number* with *amount* can be regarded as an instance of approximation: the loosening of the pattern when a verbatim association becomes approximated semantically or structurally. Both *number* and *amount* can be classified grammatically as quantifying determiners. Also, in the *Collins COBUILD Grammar Patterns* (Francis et al. 1998), out of the 39 meaning groups identified for the pattern N *of* n, both *amount* and *number* are categorised as belonging to the "Gang" group, where "[t]he noun indicates a group or quantity of people or things" and "[t]he noun group after *of* contains a plural form or plural noun which indicates the people or things involved" (Francis et al. 1998: 187).[11] In other words, the two words can fill a slot in the same grammatical construction. Therefore, substitution of *number* with *amount* can be regarded as variation within a grammatical class. This is what colligation permits in a unit of meaning. At the same time, semantically *number* and *amount* are very close, too. *Oxford Dictionary of English* (2010) defines *amount* as "a quantity of something" and *number* as "a quantity or amount" it its second sense. For this reason it is possible to classify the variation between them as variation within a semantic set, which fits the category of semantic preference.

11 In addition to this meaning group, *amount* in the same sense also belongs to the "Piece" group, where "[t]he noun indicates a quantity of a substance or thing" and occurs with an uncount noun (Francis et al. 1998: 186).

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM

Let us now move on to the second example. Nora has a preference for *the connection between* X *and* Y, which occurs six times in her Production Corpus and zero times in her Exposure Corpus, while in the Exposure Corpus *the relationship between* X *and* Y is preferred, occurring 27 times in exactly the same contexts (to compare, Nora uses this alternative expression only twice). Example (6) shows all six instances of the concgram CONNECTION/BETWEEN in Nora's drafts in the order of their occurrence.

(6) (Nora, Production Corpus)

1 The internet has changed **the connection between** music and place, ... ([ref], 2002b, p. 89, 90)
2 especially concerning **the connection between** music and place ([ref], 2002b, pp. 87).
3 argument here is that **the connection between** style, musical taste, and identity has grown
4 Studies have shown that a **connection between** online and offline exists for scene members.
5 the main focus should lay on **the connection between** the extravagant and the everyday life
6 a central aspect of this study is **the connection between** place and space.

In its first occurrence, the replacement of *relationship* with *connection* is probably a conscious choice since the reference at the end of the sentence and omission of quotation marks suggest that it is a paraphrase of the cited author's words. A look at the cited article reveals that there is no particular sentence Nora is paraphrasing here, but the author of the article continuously talks about *the relationship* (rather than *the connection*) *between*, mostly between music and place (example [7]).

(7) (Nora, the article from the Exposure Corpus she cited in example [6], line 1)

1 ions among music fans of **the relationship between** music and place. My empirical
2 for our understanding of **the relationship between** music and place. Using the c
3 ban mythscapes Studies of **the relationship between** music and place have concentr
4 musicians (Shank, 1994), **the relationship between** place, music and ethnic identity
5 on that constructions of **the relationship between** music and place are closely b
6 conceive of and discuss **the relationship between** music and place according to
7 of understanding **the** changing **relationship between** social and physical landscape
8 ew ways of thinking about **the relationship between** space and place is of consid
9 as a means of mapping out **the relationship between** social and geographical land
10 et, such constructions of **the relationship between** music and place are no longe
11 to our understanding of **the relationship between** music and place by considerin
12 yth-infused picturings of **the relationship between** music and the local, is bein

While it would not have been absolutely necessary for her to paraphrase the word *relationship*, it is very likely that this is what she did on purpose to avoid being accused of plagiarism. For this reason it is difficult to claim that it is an

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM

example of cognitive approximation. What is more important for us here is that Nora seems to adopt the phrase in its exact form (*the connection between* X *and* Y) and use it invariably in other contexts, too. This is especially notable since *the relationship between* X *and* Y is the pattern other authors (six different authors) in Nora's Exposure Corpus have a preference for (example [8]):

(8) (Nora, Exposure Corpus, instances presented in example [7] are not shown)

13 ompted me to think about **the relationship between** music and the built environment
14 eet Studios suggest about **the relationship between** music making and the built urban environment
15 there would not be **the** same **relationship between** services."9 The building thus
16 helped to illustrate how **the relationship between** music making and the built environment
17 ons. The article examined **the relationship between** music making and the urban environment
18 ustrate **the** complex, shifting **relationship between** popular music making and the built environment
19 and rigorous research on **the relationship between** music and the urban environment
20 tive online journals, for **the relationship between** individuals and groups on the internet
21 earch interests focus on **the relationship between** different forms of communications
22 o say that **the** unequal power **relationship between** the researcher and researched
23 or Neo-tribes? Rethinking **the relationship between** Youth, Style and Musical Taste
24 ris, K. (2000) '"Roots"? **The relationship between** the Global and the Local with
25 discuss below, however, **the relationship between** the underground and the market
26 hers have begun to study **the relationship between** the internet and subcultures.
27 or neo-tribes? Rethinking **the relationship between** youth, style and musical taste

There are no occurrences of *the connection between* X *and* Y in Nora's Exposure Corpus.

The preference Nora exhibits can be regarded as an example of the second process, the one I call *fixing*, which is the reverse to approximation: the pattern becomes more fixed which is evident in a more abstract component of a unit of meaning becoming verbatim. In this case, we can postulate a semantic preference for a semantic set of words to which *connection* and *relationship* belong which turns into a verbatim association with the particular member from this set – *connection*.

To be fair, it needs to be mentioned that Nora uses *the relationship between* herself twice (example [9]):

(9)

1 there is still a strong **relationship between** identity construction, style, and above all musical taste.
2 Translocality describes **the relationship between** "localized lifeworlds in faraway sites." ([ref.] 2002, p. 133).

The second instance again occurs in a paraphrase, with the original wording in the cited article being: *Translocality refers to the dynamic between localized lifeworlds in faraway sites*.

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM

While Nora's preference is clearly for *the connection between*, i.e., the balance of the exemplar representation of the unit is tipped in favour of this wording, other factors at play at a particular moment always have a potential to influence the eventually produced pattern: "the act of playing the game has a way of changing the rules" (Gleick 1987: 24).[12]

But let as have a look at another very clear example of fixing. Nora's concgram THERE/EXISTS has no corresponding concgrams in her Exposure Corpus. In her own writing, it occurs six times (example [10]).

(10) (Nora, Production Corpus)

1 ([ref.], 2007, pp. 2, 5). Even though **there exists** discourse between the different scenes they are
2 2.2 Metal related research **There exists** some **research on** Heavy Metal in general. But in
3 to an even smaller body of **research** than **there exists on** Heavy Metal ([ref.], 2004b, pp. 110).
4 be specialized for black metal but if **there exists** a lack of such forums generalized (extreme) metal
5 extreme scenes of Extreme Metal, **there** still **exists** discourse between the different subgenre-scenes.
6 the literature review shows that **there** already **exists** some interesting and valuable **research on** the

The pattern does not seem to be limited to the two co-occurring words only, *there* and *exists*, which merely constitute the fixed core of a more extended unit. In three out of six occurrences of the concgram, it is used to talk about *research on* X. Moreover, the author seems to lament a certain lack of research presumably in order to establish a niche for her own study which is a common move in research articles (Swales 1990). This can be postulated as a communicative purpose of the unit. There is also concession present quite clearly: for example, even though the author is prepared to admit that there is some research on X already, there is still little of it or there needs to be more (as in lines 2 and 6) or that there is "discourse between different scenes", they are still "distinct from each other" (as in line 1), and the other way round (as in line 5). Extended view of these concordance lines (see the appendix) confirms this interpretation.

Further exploration reveals that in addition to *there exists*, also *existing* co-occurs with *research*, and always in the context of a certain lack or gap that

12 A choice of a language item is a product of a complex system which works non-linearly. So just as in the game of ice hockey (which has nothing to do with the classical notion of a game and its rules often exploited in linguistics with reference to Wittgenstein) Gleick refers to in his famous dictum, acceleration of a hockey puck depends on friction, but one "cannot assign a constant importance to friction, because its importance depends on speed" which "in turn, depends on friction," one's choice of a language item to produce depends on a variety of factors, each in turn influencing the other, and is not constant.

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM

needs to be closed, as can be seen from the concordance lines for the concgram EXIST*/RESEARCH* in example (11).[13] Curiously the lemma EXIST co-occurs with *research* only in the form *existing* or in the phrase *there exists*.

(11) (Nora, Production Corpus)

1 or gender-based perspective. Often the **existing research** is conducted for Master theses e.g.[14]
2 research has been conducted and the older **existing research** is often lacking basic scientific
3 the literature review shows that **there** already **exist**s some interesting and valuable **research** on the
4 2.2 Metal related research **There exists** some **research** on Heavy Metal in general. But in
5 music marketing. Overall the **research** can **close existing gaps** as well as enrich existing studies with
6 to an even smaller body of **research** than **there exists** on Heavy Metal ([ref.], 2004b, pp. 110).
7 transformations. This **research** will **close** an **exis**ting **gap** in the studies of Extreme Metal as well as
8 to them. This **research** will therefore **close** an **existing gap** in the studies of extreme metal as well as

There are two occurrences of a similar pattern in Nora's priming language:

(12) (Nora, Exposure Corpus)

1 which has been somewhat overlooked by **existing research**. By focusing on such a transition,
2 of new media may occasion genre evolution" [6]. **Existing research** on genre suggests that media impacts

Taking everything into account, we can postulate a developing unit of meaning with the core EXIST*/RESEARCH* and the communicative purpose (semantic prosody) of establishing a niche for a study by pointing to the lack of previous research or gap in it.[15] The core collocates with *there* and CLOSE *existing gap**. It has a semantic preference for "lack" or "small quantity," expressed by such words as: *lacking, some, even smaller body of, gap*. It is also possible to say that it colligates with the existential *there* construction, but I think it is important to bear in mind that there can be an actual verbatim association with the word *there* instead of or in addition to this, more abstracted, structural association. It turns out that *research* also occurs in the *there*-construction without the participation of EXIST, but just once and immediately after a sentence where *there* + *exist* construction was already used, so it is probably a conscious paraphrase in order to avoid

13 An asterisk stands for zero or more characters, here, for example, for any suffix if there is one.

14 Extended co-text: "Black Metal has been mainly studied from a religious or gender-based perspective. Often the existing research is conducted for Master theses (e.g. [ref.]) or doctoral dissertations [...]. Still it can be seen that no studies have been done on the research problem posed here."

15 See Vetchinnikova (2014) for this interpretation of semantic prosody and its explanation.

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM

repetition. Interestingly, the communicative function in this one instance is still "gap in research" (see example [13]).

(13) (Nora, Production Corpus)
1 2007, pp. 42). This is essential because **there** is very little **research** on Black Metal in

It is clear that in none of the examples presented in this section we can meaningfully talk of operation on the open-choice principle. In the first two cases (AMOUNT/DEATHS and CONNECTION/BETWEEN), changes introduced in MWUs have not led to meaning-shift and dissolution of the unit. The writers also consistently demonstrated a preference for a particular variant form, which points to the process of fixing of the unit in their own language use. In the third case (EXIST*/RESEARCH*), a unit of meaning is developing, again through the process of fixing, where different components of the unit fuse together to form a holistic unit of meaning which the user produces on the idiom principle. Yet, even in this case, there is no evidence of creativity as the emergent patterning is not only similar to the two examples which were found in the student's Exposure Corpus but is also very much in line with patterns and conventions of academic writing, apparent, for example, in its association with a common move in research articles identified by Swales (1990). To sum up, all the changes introduced into MWUs by L2 users can be explained as oscillations towards fixing or approximation inside a unit of meaning. The units have been produced holistically on the idiom principle, even though some of their components may have become more fixed while some may have loosened.

7 Conclusion

In this paper I have argued that there are two distinct explanations of the differences between NS and ELF usage in ELF research which I have exemplified using the case of phraseological patterning. I have also attempted to show that these explanations can be traced back to different paradigms of linguistic scholarship: classical Saussurean and usage-based. This way, the different explanations of ELF forms, as creative actualisations of the possible and as natural variation arising due to cognitive constraints of multilingual processing, are consistent with basic assumptions of their respective paradigms.

In the empirical part of the paper I have attempted to contribute to the discussion of the specifics of phraseological patterning in ELF by adopting a

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM

usage-based perspective and (i) providing quantitative evidence on the extent to which MWUs used by ELF speakers in their writing are novel versus recycled from previous experience with the language; (ii) analysing non-matching patterns qualitatively to explore in which way they can be considered novel combinations versus based on previous exposure by applying the model of a unit of meaning which permits variability without requiring a switch to the open-choice principle. With regard to the first question, the comparison of language L2 users produce and the language they were exposed to revealed remarkable continuity: 56% to 75% of significant patterns extracted from L2 usage corpora matched the patterning of the priming language. When the same patterns were compared to the patterning of the language they were *not* exposed to, i.e., to the language from a different field, the percentages dropped to 20%–31%. The qualitative analysis of the non-matching patterns showed that apparent discrepancies can arise due to loosening or fixing of the components inside a unit of meaning. These two processes can be viewed as aspects of operation on the idiom principle. In other words, what might appear to be novel patterning at first sight seems to be a result of natural variation permissible in a unit of meaning rather than a product of creative open-choice assembling.

The reason why it is important to establish whether not quite nativelike phraseological patterning in ELF is a result of creation or approximation and whether on the whole ELF speakers predominantly operate on the open-choice or on the idiom principle is that the answers to these questions have far-reaching implications for our understanding of the properties of L2 processing and use, which in turn must have long-term consequences for language change. It is perhaps the most commonly stated fact in the literature on global English and ELF in particular that there are now more NNSs of English than NSs. Thus, if NNSs process language in a fundamentally different way from NSs, and the majority of their lexical choices are creative and therefore arbitrary rather than conditioned by previous exposure and therefore predictable, the consequences of this for the trajectory of language change must be profound.[16]

16 It must be noted, though, that this view on language change, no matter which scenario is accepted, is consistent with a usage-based perspective on language which permits modelling language as a complex adaptive system. How does English stay recognisably English? In accordance with the perspective maintaining a classical position, it is the virtual language, or the existence of a shared code all speakers of the language draw on, which provides the common ground for all ELF and ENL realisations of the encoding possibilities and so ensures the coherence of language as a whole for English to remain English (see Seidlhofer 2011: 110–112). In complex adaptive system modelling, no such common ground or design is needed since coherence in the face of change is a natural property of all complex systems.

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM

Appendix

1. *Even though there exists discourse between the different scenes they are distinct from each other and often the differentiation is essential to members.* (concession)
2. *There exists some research on Heavy Metal in general. But in comparison with research on other musical subculture, one finds very confined data concerning metal audiences.* ("lack")
3. *The general obscurity lead to an even smaller body of research than there exists on Heavy Metal.* ("lack")
4. *Since the data collection will be targeted, the forums preferably will be specialized for black metal but if there exists a lack of such forums generalized (extreme) metal forums with black metal subgroups might be included as well.* ("lack")
5. *It is eminent to note is that even though Black Metal is one of the most extreme scenes of Extreme Metal, there still exists discourse between the different subgenre-scenes. Therefore many of the assumptions and conclusions made about Extreme Metal can be adopted for the Black Metal scene ([ref.], 2007, p. 42). This is essential because there is very little research on Black Metal in particular.* (concession, though cf. example [1] above)
6. *In conclusion the literature review shows that there already exists some interesting and valuable research on the different aspects of musical subcultures and media. It also demonstrates that concerning Heavy Metal and Extreme Metal (and Black Metal) little research has been conducted and the older existing research is often lacking basic scientific standards.* ("exists but still lacking")

References

Aarts, Bas. 2001. Corpus linguistics, Chomsky and fuzzy tree structures. In Christian Mair & Marianne Hundt (eds.), *Corpus linguistics and linguistic theory*, 5–13. Amsterdam: Rodopi.

Altenberg, Bengt. 1998. On the phraseology of spoken English: The evidence of recurrent word combinations. In Anthony P. Cowie (ed.), *Phraseology: Theory, analysis and applications*, 101–122. Oxford: Clarendon.

Beckner, Clay, Richard Blythe, Joan Bybee, Morten H. Christiansen, William Croft, Nick C. Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman & Tom Schoenemann. 2009. Language is a complex adaptive system: Position paper. *Language learning* 59(s1). 1–26.

Bolinger, Dwight. 1976. Meaning and memory. *Forum Linguisticum* 1(1). 1–14.

Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82(4). 711–733.

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM

Bybee, Joan. 2008. Usage-based grammar and second language acquisition. In Peter Robinson & Nick C. Ellis (eds.), *Handbook of cognitive linguistics and second language acquisition*, 216–236. London: Routledge.

Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.

Bybee, Joan & Paul Hopper. 2001a. *Frequency and the emergence of linguistic structure*. Amsterdam & Philadelphia: John Benjamins.

Bybee, Joan & Paul Hopper. 2001b. Introduction to frequency and the emergence of linguistic structure. In Joan L. Bybee & Paul Hopper (eds.), *Frequency and the emergence of linguistic structure*, 1–24. Amsterdam & Philadelphia: John Benjamins.

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

Carter, Ronald & Michael McCarthy. 2006. *Cambridge grammar of English: A comprehensive guide*. Cambridge: Cambridge University Press.

Cheng, Winnie, Chris Greaves & Martin Warren. 2006. From n-gram to skipgram to concgram. *International Journal of Corpus Linguistics* 11(4). 411–433.

Cheng, Winnie, Chris Greaves, John McH. Sinclair & Martin Warren. 2009. Uncovering the extent of the phraseological tendency: Towards a systematic analysis of concgrams. *Applied Linguistics* 30(2). 236–252.

Chomsky, Noam. 1986. *Knowledge of language: Its nature, origin, and use*. Westport, CT: Greenwood Publishing.

Croft, William A. 2001. *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.

Dąbrowska, Eva. 2004. *Language, mind and brain*. Edinburgh: Edinburgh University Press.

ELFA corpus. The corpus of English as a lingua franca in academic settings. Director: Anna Mauranen. http://www.helsinki.fi/elfa/elfacorpus (accessed 7 July 2015).

Ellis, Nick C. 2011. The emergence of language as a complex adaptive system. In James Simpson (ed.), *The Routledge handbook of applied linguistics*, 666–679. London: Routledge.

Ellis, Nick C. 2012. What can we count in language, and what counts in language acquisition, cognition, and use? In Stefan Th. Gries & Dagmar S. Divjak (eds.), *Frequency effects in language learning and processing (vol. 1)*, 7–34. Berlin & New York: Mouton de Gruyter.

Ellis, Nick C. & Diane Larsen-Freeman. 2006. Language emergence: Implications for Applied Linguistics. [Introduction to the special issue]. *Applied Linguistics* 27(4). 558–589.

Erman, Britt & Beatrice Warren. 2000. The idiom principle and the open choice principle. *Text* 20(1). 29–62.

Francis, Gill, Susan Hunston & Elizabeth Manning. 1998. *Collins COBUILD Grammar Patterns: Nouns and adjectives*. London: HarperCollins.

Gleick, James. 1987. *Chaos: Making a new science*. New York: Penguin.

Goldberg, Adele. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Goldberg, Adele. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Granger, Sylviane. 1998. Prefabricated patterns in advanced EFL writing: Collocations and formulae. In Anthony P. Cowie (ed.), *Phraseology: Theory, analysis, and applications*, 145–160. Oxford: Clarendon.

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM

Granger, Sylviane & Fanny Meunier (eds.). 2008. *Phraseology. An interdisciplinary perspective*. Amsterdam & Philadelphia: John Benjamins.

Greaves, Chris. 2009. *ConcGram 1.0: A phraseological search engine. [Software]*. Amsterdam: John Benjamins.

Gurevich, Olga, Matthew A. Johnson & Adele E. Goldberg. 2010. Incidental verbatim memory for language. *Language and Cognition* 2(1). 45–78.

Hopper, Paul. 1987. Emergent grammar. *Berkeley Linguistics Society* 13. 139–157.

Howarth, Peter. 1998. The phraseology of learners' academic writing. In Anthony P. Cowie (ed.), *Phraseology: Theory, analysis, and applications*, 161–186. Oxford: Clarendon.

Hunston, Susan. 2002. *Corpora in applied linguistics*. Cambridge: Cambridge University Press.

Hynninen, Niina. 2014. The Common European Framework of Reference from the perspective of English as a lingua franca: What we can learn from a focus on language regulation. *Journal of English as a Lingua Franca* 3(2). 291–314.

Jenkins, Jennifer. 2000. *The phonology of English as an international language*. Oxford: Oxford University Press.

Kecskes, Istvan. 2007. Formulaic language in English lingua franca. In Istvan Kecskes & Laurence Horn (eds.), *Explorations in pragmatics: Linguistic, cognitive and intercultural aspects*, 191–218. Berlin & New York: Mouton de Gruyter.

Larsen-Freeman, Diane. 2013. Complexity theory/dynamic systems theory. In Peter Robinson (ed.), *The Routledge encyclopaedia of second language acquisition*, 103–106. New York: Routledge.

Laury, Ritva & Tsuyoshi Ono. 2005. Data is data and model is model: You don't discard the data that doesn't fit your model! *Language* 81(1). 218–225.

Mauranen, Anna. 2005. English as a lingua franca – An unknown language? In Giuseppina Cortese & Anna Duszak (eds.), *Identity, community, discourse: English in intercultural settings*, 269–293. Frankfurt: Peter Lang.

Mauranen, Anna. 2009. Chunking in ELF: Expressions for managing interaction. *Journal of Intercultural Pragmatics* 6(2). 217–233.

Mauranen, Anna. 2011. Learners and users – Who do we want corpus data from? In Fanny Meunier, Sylvie De Cock, Gaëtanelle Gilquin & Magali Paquot (eds.), *A taste for corpora: In honour of Sylviane Granger*, 155–171. Amsterdam & Philadelphia: John Benjamins.

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

McEnery, Tony & Andrew Hardie. 2013. The history of corpus linguistics. In Keith Allan (ed.), *The Oxford handbook of the history of linguistics*, 727–745. Oxford: Oxford University Press

McEnery, Tony, Richard Xiao & Yukio Tono. 2006. *Corpus-based language studies: An advanced resource book*. London & New York: Routledge.

Meunier, Fanny & Sylviane Granger (eds.). 2008. *Phraseology in foreign language learning and teaching*. Amsterdam & Philadelphia: John Benjamins.

Meyer, Charles F. & Hongyin Tao. 2005. Response to Newmeyer's "Grammar is grammar and usage is usage." *Language* 81(1). 226–228.

Nesselhauf, Nadja. 2005. *Collocations in a learner corpus*. Amsterdam & Philadelphia: John Benjamins.

Newmeyer, Frederick J. 2003. Grammar is grammar and usage is usage. *Language* 79(4). 682–707.

Oxford Dictionary of English (ODE). 2010. 3rd edn. Oxford: Oxford University Press.

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM

Partington, Alan. 2006. *The linguistics of laughter: A corpus-assisted study of laughter talk*. London & New York: Routledge.
Pawley, Andrew & Frances H. Syder. 1983. Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In Jack C. Richards & Richard W. Schmidt (eds.), *Language and communication*, 191–227. London: Longman.
Pitzl, Marie-Luise. 2009. "We should not wake up any dogs": Idiom and metaphor in ELF. In Anna Mauranen & Elina Ranta (eds.), *English as a lingua franca: Studies and findings*, 298–322. Newcastle upon Tyne: Cambridge Scholars Press.
Pitzl, Marie-Luise. 2012. Creativity meets convention: Idiom variation and re-metaphorization in ELF. *Journal of English as a Lingua Franca* 1(1). 27–55.
Schmitt, Norbert (ed.). 2004. *Formulaic sequences: Acquisition, processing and use*. Amsterdam & Philadelphia: John Benjamins
Searle, John R. 1969. *Speech acts*. Cambridge: Cambridge University Press.
Searle, John R. 1995. *The construction of social reality*. New York: Free Press.
Seidlhofer, Barbara. 2009. Accommodation and the idiom principle in English as a Lingua Franca. *Journal of Intercultural Pragmatics* 6(2). 195–215.
Seidlhofer, Barbara. 2011. *Understanding English as a Lingua Franca*. Oxford: Oxford University Press.
Seidlhofer, Barbara & Henry G. Widdowson. 2007. Idiomatic variation and change in English. The idiom principle and its realizations. In Ute Smit, Julia Hüttner, Gunther Kaltenböck & Ursula Lutzky (eds.), *Tracing English through time: Explorations in language variation*, 359–374. Vienna: Braumüller.
Sinclair, John McH. 1987. Collocation: A progress report. In Ross Steele & Terry Treadgold (eds.), *Language topics: Essays in honour of Michael Halliday*, 319–331. Amsterdam: John Benjamins.
Sinclair, John McH. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
Sinclair, John McH. 1996. The search for units of meaning. *Textus* 9(1). 75–106.
Sinclair, John McH. 2004. *Trust the text*. London: Routledge.
Stubbs, Michael. 2001. Texts, corpora, and problems of interpretation: A response to Widdowson. *Applied Linguistics* 22(2). 149–172.
Swales, John. 1990. *Genre analysis: English in academic and research settings*. New York: Cambridge University Press.
Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
van Gelderen, Elly. 2011. Grammaticalization and generative grammar: A difficult liaison. In Heiko Narrog & Bernd Heine (eds.), *The Oxford handbook of grammaticalization*, 43–55. Oxford: Oxford University Press.
Vetchinnikova, Svetlana. 2014. *Second language lexis and the idiom principle*. Helsinki: University of Helsinki doctoral dissertation. http://urn.fi/URN:ISBN:978-951-51-0064-1 (accessed 22 April 2015).
Widdowson, Henry G. 1997. EIL, ESL, EFL: Global issues and local interests. *World Englishes* 16(1). 135–146.
Widdowson, Henry G. 2000. On the limitations of linguistics applied. *Applied Linguistics* 21(1). 3–25.
Widdowson, Henry G. 2003. *Defining issues in English language teaching*. Oxford: Oxford University.

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM

Widdowson, Henry G. 2012. ELF and the inconvenience of established concepts. *Journal of English as a Lingua Franca* 1(1). 5–26.
Wray, Alison. 2002. *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
Yorio, Carlos A. 1989. Idiomaticity as an indicator of second language proficiency. In Kenneth Hyltenstam & Loraine K. Obler (eds.), *Bilingualism across the lifespan*, 55–72. Cambridge: Cambridge University Press.

Bionote

Svetlana Vetchinnikova

Svetlana Vetchinnikova is a post-doctoral researcher in the Changing English project at the University of Helsinki. She has recently defended her doctoral thesis entitled *Second Language Lexis and the Idiom Principle* (2014). Her current research centres on language patterning at different planes of its organisation and the underlying processing mechanisms, especially in second language users.

Brought to you by | The Helsinki University Library
Authenticated
Download Date | 9/24/15 3:41 PM