

Istvan Kecskes*

State University of New York, Albany
USA

THE IDIOM PRINCIPLE IN ENGLISH AS A LINGUA FRANCA

Abstract

This paper examines how the “idiom principle” (Sinclair 1991) works in the language use of English as a Lingua Franca speakers. It is hypothesized that the idiom principle that drives word selection in monolinguals may be blocked in the L2 of bilinguals and the “open choice principle” governs instead. In order to investigate the validity of this hypothesis a small corpus of non-native speaker – non-native speaker (lingua franca) communication is examined and compared to a similar study (Kecskes 2007) where the bilingual speakers used their L2 (English).

Based on the two studies we can conclude that the “idiom principle” is the most salient guiding mechanism in any language production. But it results in less formulaic language use in L2 than in L1 of bilinguals. This claim basically concurs with the findings of other studies (cf. Bolander 1989; Pawley and Syder 1983; Warga 2008; Weinert 1995) that also talked about the limited use of formulaic language in L2.¹

Keywords: idiom principle, formulaic language, psychological salience, second language idiom processing, lingua franca, open choice principle

* E-mail address: ikecskes@albany.edu

¹ This is a version of a study that was published as Chapter 2. in Roberto Heredia and Anna Cieslicka (eds.) *Bilingual Figurative Language Processing*. Cambridge: Cambridge University Press. Pp. 28-53, 2015.

1. Introduction

Based on the findings of some studies that argued that even advanced L2 users have difficulty with formulaic language (Ellis, Simpson-Vlach, & Carson, 2008; Kecskes, 2007; Prodromou 2008; Warga, 2005) this paper hypothesizes that the *idiom principle* (Sinclair, 1991) that drives word selection in monolinguals may be blocked in the L2 of bilinguals and the *open choice principle* governs instead. In order to investigate the validity of this hypothesis a small corpus of non-native speaker – non-native speaker (*lingua franca*) communication is examined where the bilingual speakers use their L2 (English).

Before analyzing bilingual speech production there is a need to discuss utterance production in general and discuss why formulaic language has a special role in both production and comprehension. In some recent publications (Kecskes, 2012, 2014), I argued that speaker's utterance, no matter what language s/he uses is the result of the interplay of salience and recipient design. While fitting words into actual situational contexts, speakers are driven not only by the intent (*conscious*) that the hearer recognize what is meant as intended by the speaker (*cooperation*), but also by speaker individual salience that affects production subconsciously (*egocentrism*). The interplay of these social (*cooperation*) and individual factors (*egocentrism*) shapes the communicative process.

In order to succeed, speakers must correctly express intended illocutionary acts by using appropriate words, and make their attempt in an adequate context. Speakers relate propositional contents to the world (actual situational context; audience) with the intention of establishing a correspondence between words and things from a certain direction of fit. Speaker's utterance is a full proposition that is the result of the speaker's intention that is a private reaction to a communicative situation. Speaker's intention is expressed in lexical items whose selection is affected not only by recipient design but also by speaker's egocentrism governed by salience. Salience, which operates subconsciously and automatically, may affect word selection and utterance formation just like recipient design.

Salience affecting language production is motivated by both private and collective elements. What is salient for a speaker is based on prior experience of that individual. However, the mechanism according to which salience works is also affected by universal elements and forces that language users share with others. One such phenomenon is *the economy*

principle in language use. Human beings want to achieve as much as possible with the least possible effort both in production and comprehension. In the *Relevance Theory* (Sperber & Wilson, 1995) *economy* has been used with two functions: first to explain how cognitive processes are linked to utterance interpretation (processing efforts must be balanced by cognitive effects), second to explain how communication may be successful (inferences complete the underspecified content of the utterance to obtain its intended meaning). In other words, economy is a property of the cognitive system devoted to utterance interpretation, and is also required in order to ensure successful communication, by the computational devices, which are combined with linguistic decoding to yield the intended meaning of an utterance.

There is psycholinguistic evidence that fixed expressions and formulas have an important economizing role in speech production (cf. Miller & Weinert, 1998; Wray, 2002). Sinclair's (1991) *idiom principle* says that the use of prefabricated chunks may ... *illustrate a natural tendency to economy of effort* (1991: 110). This means that in communication we want to achieve more cognitive effects with less processing effort. Formulaic expressions ease the processing overload not only because they are *ready-made* and do not require of the speaker/hearer any *putting together*, but also because their salient meanings are easily accessible in online production and processing.

Wray (2002) said that by favoring formulaic units speakers are able to reduce both their own processing – the larger the units, the fewer the operations needed to construct the message – and also the processing load of the hearer. She argued that there are major benefits to the speaker in ensuring that the hearer does not have to engage in too much processing. She also proposed that both parties are to some extent obliged to find ways of minimizing their processing, because the grammar of human language is too complex for human memory to cope with all the time (Wray, 2002: 15). Thus Wray converged with Sinclair's proposal (1991) that the formulaic option, which he calls the *idiom principle*, is the *default processing strategy*. Analytic processing, the *open choice principle* in Sinclair's terminology, is invoked only when the idiom principle fails.

This is the main point in this paper. *Being the default processing strategy the formulaic option (i.e., idiom principle) is expected to be most salient in language production*. It means that the speaker is expected to come up primarily with utterances that contain ready-made, formulaic expression(s)

if possible and plausible. If it is not, the *open choice principle* steps in. This looks like a logical mechanism in monolingual language production where participants can rely on the mutual understanding of formulaic expressions that are motivated by common ground, conventions, commonalities, norms, common beliefs and mutual knowledge. But is that also the case in bilingual speakers when they use their L2? Does their mind prewired for the *idiom principle* follow pass with the required circumstances (partly) missing? No matter which of their languages they use, bilingual speakers, to some extent, miss common ground, conventions, commonalities, norms, etc. In order for us to answer these questions, we first need to look at how the *idiom principle* works in monolingual language production.

2. Formulaic Units

By *formulaic language* we usually mean multi-word collocations which are stored and retrieved holistically rather than being generated *de novo* with each use. Collocations, fixed semantic units, frozen metaphors, phrasal verbs, speech formulas, idioms and situation-bound utterances can all be considered as examples of formulaic language (Howarth, 1998; Kecskes, 2000; Wray, 1999, 2005, 2005). These word strings occurring together tend to convey holistic meanings that are either more than the sum of the individual parts, or else diverge significantly from a literal, or word-for-word meaning and operate as a single semantic unit (Gairns & Redman, 1986: 35).

Formulaic language is the heart and soul of native-like language use. In intercultural communication, the type of communication in which bilinguals usually participate most frequently, one of the major issues is to decide how exactly we expect interlocutors to use the common language, the *lingua franca*, which is English in this study. Is it enough for the participants just to use the common language as a system of linguistics signs (sticking mainly to the literal meanings of lexical units) with possible meanings that are disambiguated and negotiated in the process of interaction, or do we expect that the interlocutors stick to the rules of the game and do similarly to what the native speakers of that language do (i.e., rely on both prefabricated chunks and *ad-hoc* generated elements and combine them in a creative way?).

In monolingual production there is hardly any doubt about the salience of the *idiom principle*. Coulmas (1981: 1) argued that much of what is actually said in everyday conversation is by no means unique. *Rather, a great deal of communicative activity consists of enacting routines making use of prefabricated linguistic units in a well-known and generally accepted manner.* He continued claiming that *successful co-ordination of social intercourse heavily depends on standardized ways of organizing interpersonal encounters* (p. 3). Howarth (1998) also talked about the fact that native speaker linguistic competence has a large and significant phraseological component. This means that the ability to sound idiomatic (i.e., achieving *nativelike selection*, in the words of Pawley and Syder, 1983) plays a very important role in language production and comprehension. This fact has a profound effect on how we explain intercultural interaction because both figurative and formulaic language is the result of conventionalization and standardization which is supported by regular use of certain lexical units for particular purposes in a speech community. This is usually what non-native speakers have limited access to in the target language.

People using a particular language and belonging to a particular speech community have *preferred ways of saying things* (cf. Kecses, 2007; Wray, 2002) and *preferred ways of organizing thoughts* (Kecses, 2007). Preferred ways of saying things are generally reflected in the use of formulaic language and figurative language, while preferred ways of organizing thoughts can be detected through analyzing, for instance, the use of subordinate conjunctions, clauses and discourse markers. Selecting the right words and expressions and formulating utterances in ways preferred by the native speakers of that language (*nativelike selection*) is more important than syntax. The following example from a sign in an Austrian hotel catering to skiers (Octopus, 1995: 144) demonstrates this clearly.

- (1) *Not to perambulate the corridors in the hours of repose in the boots of descension.*
Correctly: *Don't walk in the halls in ski boots at night.*

The sentence shows absolutely bad word choices but acceptable syntax. The content of the sentence can be expressed by different word combinations. However, the words selected by the writer of this sentence hardly make much sense when put together even if the right syntax is used. The person

was likely to have used a dictionary with word equivalency as a selection principle.

3. Psychological Saliency and the Formulaic Continuum

Psychological saliency of word sequences

The importance of formulaic language was noticed in earlier linguistic research. Hymes (1962) pointed out that an immense portion of verbal behavior consists of linguistic routines. Bolinger (1976, p. 2) suggested that *speakers do at least as much remembering as they do putting together*. Fillmore (1976: 4) also found that *an enormously large amount of natural language is formulaic, automatic and rehearsed, rather than propositional, creative or freely generated*. However, with the appearance of huge corpora, understanding formulaic language has become more complicated. Working with large corpora Altenberg (1998) went so far as to claim that almost 80% of our language production can be considered formulaic. Whatever the proportion actually is, one thing is for sure: speakers in conventional speech situations tend to do more remembering than putting together. *Our everyday conversations are often restricted to short routinized interchanges where we do not always mean what we say*. So a typical conversation between a customer and a store assistant may look like this:

(2) Conversation between store assistant (A), and Customer (B).

A: – *What can I do for you?*

C: – *Thank you, I am just looking.*

A: – *Are you looking for something particular?*

C: – *No, not really.*

A: – *If you need help, just let me know.*

None of the expressions used by the speakers look freely generated. Each of them can be considered a formula that is tied to this particular kind of situation. However, if we consider the following conversation, we may see something different.

(3) Sam (S) and Bob (B) are talking.

S: – ***If you want to see me again you will need to do what I tell you to.***

B: – ***OK, my friend.***

Can the expressions in bold be considered formulas? Are they in any way different from the ones in example (2)? There is no doubt that the expressions in bold consist of words that are frequently used together. But are they formulas here? Do they have some kind of psychological saliency as formulas for the speakers? We must be careful with the answer because *frequency is only one of the criteria* based on which we can identify formulaic expressions. The problem is that the role of frequency seems to be overemphasized in present day linguistics, especially in corpus linguistics. Recent research analyzing written and spoken discourse has established that highly frequent, recurrent sequences of words, variously called lexical bundles, chunks, and multiword expressions, are not only salient but also functionally significant. Cognitive research demonstrated that knowledge of these ready-made expressions is crucial for fluent processing. The recurrent nature of these units is discussed in the relevant literature (Biber, Johansson, Leech, Conrad, Finegan 1999; McEnery & Wilson, 1996). Simpson-Vlach and Ellis (2010) confirmed that large stretches of language are adequately described as collocational streams where patterns flow into each other. However, Sinclair's (1991) *idiom principle* is based not primarily on frequency that results in long lists of recurrent word sequences (Biber, Conrad & Cortes 2004; Biber et al. 1999), which hardly give any chance to distinguish where we have conventionalized formulas or where we have just frequently occurring word chunks that lack psychological saliency. Biber et al. (1999: 990), in their study of *lexical bundles*, defined formulaic language as *sequences of word forms that commonly go together in natural discourse*, irrespective of their structural make-up or idiomaticity, and argued that conversation has a larger amount of lexical bundle types than academic prose. However, there seems to be a clear difference from the perspective of psychological saliency between sequences such as *to tell the truth, as a matter of fact* on the one hand, and *if they could ...* or *to make it* on the other, although all these expressions are high on any frequency-based list. This is why we need to distinguish between groups of prefabricated expressions that have psychological saliency for speakers of a particular language community and loosely-tied, frequently occurring word-sequences (usually consisting of common words) such as *if they want, to do with it, and of the, tell them to*, etc. Simpson-Vlach and Ellis (2010) argued that psycholinguistically salient sequences like *on the other hand, suffice it to say* cohere much more than would be expected by chance. They are *glued together* and thus measures of association, rather than raw frequency, are likely more relevant to these formulaic expressions.

Second language (L2) studies that are relevant for bilinguals show something different. They emphasize the importance of frequency in processing formulaic language. Ellis et al. (2008) argued that formula processing by nonnatives, despite their many years of English as a second language (ESL) instruction, was a result of the frequency of the string rather than its coherence. For learners at that stage of development, it is the number of times the string appears in the input that determines fluency. Ellis et al. argued that tuning the system according to frequency of occurrence alone is not enough for nativelike accuracy and efficiency. According to those authors, what is additionally required is tuning the system for coherence—for co-occurrence greater than chance. Ellis et al. (2008) claimed that this is what solves the two puzzles for linguistic theory posed by Pawley and Syder (1983), nativelike selection and nativelike fluency. Native speakers have extracted the underlying co-occurrence information, often implicitly from usage; nonnatives, even advanced ESL learners with more than ten years of English instruction, still have a long way to go in their sampling of language. These learners are starting to recognize and become attuned to more frequent word sequences, but they need help to recognize distinctive formulas.

Why is this issue important for bilinguals? It is because the development of psychological validity/saliency of these expressions in L2 is *a matter of not only frequency and exposure to the language use but also immersion in the culture and the wish of the nonnative speaker whether s/he wants to use them or not*. Frequent encounters with these expressions for nonnative speakers help but are not enough to develop psychological saliency, as the following encounter between a Korean student and a clerk at the Registrar's office demonstrates:

(4) Korean student (Lee) and Registrar (Clerk) encounter.

Lee: – *Could you sign this document for me, please?*

Clerk: – ***Come again...?***

Lee: – *Why should I come again? I am here now.*

In spite of the distinctive intonation used by the clerk when uttering *come again*, the Korean student processed the expression not as a formula but a freely generated expression with literal meaning. So what really counts is the *measures of association, rather than raw frequency*. *What creates psychological saliency is the discursive function in a particular context of that expression*. The functional aspect is what makes immersion in the culture

important for nonnative speakers, because that is where those functions come from.

The difference in developing and using formulaic language in native and nonnative speakers raises the questions: *Not having “nativelike selections” skills and “nativelike fluency” how much can bilingual speakers stick to the original rules of the game in intercultural interactions when using their L2? How salient is the “idiom principle” in L2 language production of lingua franca speakers?*

I will try to answer the question by analyzing natural language data. However, before the analysis we will briefly need to have a look at the various groups of formulas that will be subject to analysis.

The formulaic continuum

Certain language sequences have conventionalized meanings which are used in predictable situations. This functional aspect, however, is different in nature in each type of fixed expression, which justifies the hypothesis of a *continuum* (Kecses 2003, 2007) that contains grammatical units (e.g. *be going to*) on the left, fixed semantic units (cf. *as a matter of fact*; *suffice it to say*) in the middle, and pragmatic expressions (such as situation-bound utterances: *welcome aboard*; *help yourself*) and idioms (*make ends meet*, *spill the beans*) on the right. This continuum (see Table 1 below) categorizes only those expressions that are motivated and have some psychological saliency for the speakers of a speech community.

Table 1. Formulaic Continuum

Grammar Units	Fixed Semantic Units	Phrasal Verbs	Speech Formulas	Situation-bound Utterances	Idioms
<i>going to</i>	<i>as a matter of fact</i>	<i>put up with</i>	<i>going shopping</i>	<i>welcome aboard</i>	<i>kick the bucket</i>
<i>have to</i>	<i>suffice it to say</i>	<i>get along with</i>	<i>not bad you know</i>	<i>help yourself</i>	<i>spill the beans</i>

The more we move to the right on the functional continuum, the wider the gap seems to become between compositional meaning and actual situational meaning of expressions. Language development often results in a change of function (i.e. a right to left or left to right movement of a linguistic unit on the continuum). Lexical items such as *going to* can become grammaticalized, or lexical phrases may lose their compositionality and develop an *institutionalized* function, such as *I'll talk to you later*, *Have a nice day*, *Welcome aboard*, *Be my guest*, and the like. Speech formulas such as *you know*, *not bad*, *that's all right* are similar to situation-bound utterances (SBU). The difference between them is that while SBUs are usually tied to particular speech situations, speech formulas can be used anywhere in the communication process where the speakers find them appropriate. See, for instance, the difference between *nice to meet you* and *you know* or *have a nice weekend* and *kinda*.

4. Pilot study

In 2007 I conducted a cross-sectional survey to investigate how bilingual English *Lingua Franca* speakers use formulaic language in order to answer the following question: With no native speakers participating in the language game how much will the players stick to the original rules of the game? (Kecskes, 2007) I thought that the best way to answer this question is to focus on formulaic expressions that are the reflections of *nativelikeness* that is best defined as knowing preferred ways of saying things and preferred ways of organizing thoughts in a language.

Data were collected in spontaneous *lingua franca* communication. Participants were 13 adult individuals in two groups with the following first languages: Spanish, Chinese, Polish, Portuguese, Czech, Telugu, Korean and Russian. All subjects came from the Albany community and had spent a minimum of six months in the U.S. and had at least intermediate knowledge of English before arriving. None of them had English as their first language. Both Group 1 (7 students) and Group 2 (6 students) participated in a 30-minute discussion about the following topics: housing in the area, jobs, and local customs. The conversations were undirected, and uncoached. Subjects said what they wanted to say. No native speaker was present. Conversations were recorded and then transcribed, which resulted in a 13,726 word database.

Data analysis focused on the types of formulaic units given in Table 1 above. The questions I sought to answer can be summarized as follows:

- (1) How does the use of formulas relate to the ad hoc generated expressions in the data?
- (2) What type of fixed expressions did the subjects prefer?
- (3) What formulas did speakers create on their own?

Findings

The database consists of 13,726 words. Table 2 below shows the number of words that represent the six types of formulaic units that I focused on in the database. Words were counted in each type of formulaic chunk in the transcripts. Following are samples for each unit:

- Grammatical units: I am **going to** stay here; you **have to** do that
- Fixed semantic units: **after a while, for the time being, once a month, for a long time**
- Phrasal verbs: they were **worried about** me; **take care of** the kids
- Speech formulas: **not bad; that's why; you know; I mean; that's fine**
- Situation-bound utterances: **how are you?; have a nice day; you are all set**
- Idioms: **give me a ride; that makes sense**

Table 2. Number of Expressions that Represent the Six Types of Formulaic Units

Grammar Units	Fixed Semantic Units	Phrasal Verbs	Speech Formulas	Situation-bound Utterances	Idioms	Total
102	235	281	250	57	115	1040

What is striking is the relatively low occurrence of formulaic expressions in the database. There were 1,040 formulas total used as formulaic expressions out of 13,726 in the corpus, which is only 7.6%. Even if we know that this low percentage refers only to one particular database, and the results may change significantly if our focus is on other databases, it is still much less

than linguists speak about when they address the issue of *formulaicity* in native speaker conversation. Even if our database is very limited and does not let us make generalizations about *lingua franca* communication, one thing seems to be obvious. *As far as formulaic language use is concerned there seems to be a significant difference between native speaker communication and lingua franca communication with bi- and multilingual speakers.* Non-native speakers appear to rely on prefabricated expressions in their *lingua franca* language production to a much smaller extent than native speakers. The question is why this is so. To give an answer to the question we should look at the distribution of formula types in the database displayed in Table 2.

Most frequent occurrences are registered in three groups: fixed semantic units, phrasal verbs, and speech formulas. It is interesting to mention that Ortactepe (2012) also found in her study that these three types of formulaic expressions are the ones most used and preferred in nonnative speaker language production. However, we have to be careful with speech formulas that constitute a unique group, because if we examine the different types of expressions within the group we can see that three expressions (*you know; I / you mean; you're right*) account for 66.8% (167 out of 250) of all units counted in this group. The kind of frequency that we see in the use of these three expressions is not comparable to any other expressions in the database. This seems to make sense because these particular speech formulas may fulfill a variety of different functions such as *back-channeling* (i.e., cases in which a listener utters short speech formulas such as *right, I see, OK, etc.* to signal to the speaker that she follows or agrees with him), *filling a gap*, and the like. They are also used very frequently by native speakers so it is easy for non-native speakers to pick them up.

If we disregard speech formulas for the reason explained above, formulas that occur in higher frequency than any other expressions are fixed semantic units and phrasal verbs. We did not have a native speaker control group, but we can speculate that this might not be so in native speaker communication. It can be hypothesized (based on studies mentioned earlier) that native speakers use the groups of formulas in a relatively balanced way, or at least in their speech production fixed semantic units and phrasal verbs do not show priority to the extent shown in *lingua franca* communication. How can this preference of fixed semantic units and phrasal verbs by non-native speakers be explained? How does this issue relate to the first observation about the amount of formulas in native speaker communication and *lingua franca* communication of bilinguals?

As the *think aloud* sessions (in which subjects talked about their own language production) demonstrated, the two issues are interrelated. English as a *Lingua Franca* (ELF) speakers usually avoid the use of formulaic expressions not necessarily because, as they explained, they do not know these phrases, but because they are worried that their interlocutors, who are also nonnative speakers, will not understand them properly. They are reluctant to use language that they know, or perceive to be figurative or semantically less transparent (see also Philip 2005). ELF speakers try to come as close to the compositional meaning of expressions as possible because they think that, if there is no figurative and/or metaphorical meaning involved, their partners will process the English words and expressions the way they meant them. Since bilingual speakers come from different socio-cultural backgrounds and represent different cultures, *the mutual knowledge they may share is usually restricted to the knowledge of the linguistic code*. Consequently, semantic analyzability seems to play a decisive role in ELF speech production. This assumption is supported by the fact that the most frequently used formulaic expressions are the fixed semantic units and phrasal verbs in which there is semantic transparency to a much greater degree than in idioms, situation-bound utterances, or speech formulas. Of course, one can argue that phrasal verbs may frequently express figurative meaning and function like idioms, such as *I never hang out...*; *they will kick me out from my home...* However, when I found cases like this in the database, I listed the phrasal verb among the category “idioms” rather than “phrasal verbs.” So the group of phrasal verbs above contains expressions in which there is usually clear semantic transparency.

Our subjects were more advanced speakers. This is important because there is a difference in formulaic language use between less and more proficient non-native speakers. Based on longitudinal studies both Howarth (1998) and Ortactepe (2011) came to the conclusion that less proficient learners pick up formulaic expressions and overuse them, while more advanced learners prefer to *generate* their own sentences rather than resorting to prefabricated units, a process that Howarth (1998, p. 29) refers to as *deliberate creativity*. Formulaic expressions provide non-native speakers with *survival phrases that achieve basic socio-interactional functions* (Wray & Perkins 2000: 23). They have automatic access to prefabricated chunks, and this eases communication especially in the early stages of language learning (cf. Nattinger & DeCarrico 1992; Wray 2002).

According to Segalowitz and Freed (2004), at later stages of language development, formulaic expressions function as a database for non-native speakers from which *learners abstract recurrent patterns, leading to the mastery of grammatical regularities* (2004: 403). Wray (2002) considers this creative tendency of advanced learners as a major problem resulting from *the production of perfectly grammatical utterances that are simply not used by native speakers* (2002: 147). This claim is in line with my finding about the language use of *lingua franca* speakers. Pawley and Syder (1983) referred to this deliberate creativity of relatively advanced L2 learners as a process of over-generating and producing grammatical, non-idiomatic utterances due to not having accumulated the native repertoire of formulaic expressions as *nativelike competence and fluency demand such idiomaticity* (Ellis 2003: 12).

The danger for *lingua franca* speakers in the use of formulaic language is that they often pick up these expressions without comprehending the socio-cultural load that they carry. This is especially true for situation-bound utterances in which it is usually the figurative meaning that is dominant rather than the literal meaning. In *lingua franca* communication, if one of the interactants does not know this figurative meaning and processes the utterance literally, misunderstanding may occur, such as in the following conversation between a Chinese and a Turkish student.

- (4) Conversation between a Chinese and a Turkish student;
Chinese: – I think Peter drank a bit too much at the party yesterday.
Turkish: – Eh, *tell me about it*. He always drinks too much.
Chinese: – When we arrived he drank beer. Then Mary gave him some vodka. Later he drank some wine. Oh, too much.
Turkish: – Why are you telling me this? I was there.
Chinese: -Yes, but you told me to tell about it.

Here the Turkish student used the expression “tell me about it” figuratively as a formula, while the Chinese student processed it literally. In order to avoid cases like this, *lingua franca* speakers stick to literal rather than figurative production. The use of semantically transparent language resulted in fewer misunderstandings and communication breakdowns than expected in my survey. This finding of my study corresponds with House’s observation about the same phenomena (House 2003).

Another example of this interesting phenomenon in the database is the *endeavor of speakers creating their own formulas*. This phenomenon

fully confirms the general priority of the idiom principle as most salient even in bilingual language production. Speakers (let them be mono- or multilinguals) make an effort to use formulas, not matter which language of theirs they use. This tendency was noted in some other studies as well (e.g. Cheng et al. 2009; Pitzl 2012).

The formulas our subjects created can be split into two categories. In the first category, we can find expressions that are used only once and demonstrate an effort to sound metaphorical. However, this endeavor is usually driven by the first language (L1) of the speaker in which there may be an equivalent expression for the given idea. For instance:

- (5) Formulas that demonstrate an effort to sound metaphorical.
it is almost skips from my thoughts
you are not very rich in communication
take a school

The other category comprises expressions that are created on the spot during the conversations and are picked up by the members of the *ad hoc* speech community. One of the participants creates or coins an expression that is needed in the discussion of a given topic. It becomes a part of the interculture being created (cf. Kecskes 2013). This unit functions like a *target language formula*, the use of which may be accepted by the participants in the given conversation, as demonstrated by the fact that other participants also pick it up and use it. However, this is just a temporary formula that may be entirely forgotten when the conversation is over. This is a typical example of how intercultural expressions are created. For instance:

- (6) Formulas created on the spot or *ad hoc*.
we connect each other very often
native American (in the sense of native speaker of American English)

Lingua franca speakers frequently coin or create their own ways of expressing themselves effectively, and the mistakes they may make will carry on in their speech, even though the correct form is there for them to imitate. For instance, several participants adopted the phrase *native Americans* to refer to native speakers of American English. Although in the “think aloud” conversation session, the correct expression (*native speaker of American English*) was repeated several times by one of the researchers, the erroneous formula “native Americans” kept being used by the *lingua franca* speakers. They even joked about it and said that the use of target language formulas coined by them in their temporary speech community

was considered like a “joint venture” and created a special feeling of camaraderie in the group.

Based on this study we can say that with no native speakers participating in the language game the *lingua franca* (L2) speakers can't always keep the original rules of the game. So the “*idiom principle*” does not seem to be working as it does in L1. Kecskes (2007) argued that actual speech situations in *lingua franca* communication can be considered open social situations which do not encourage the use of formulaic language. In first language communication we have much more closed social situations defined by the parameters and values taken for granted in them (see Clark 1996: 297). The result of these closed social situations is a highly routine procedure. For instance:

(7) Close social situations.

Bar: – *Two vodka tonics.*

Museum ticket booth: – *Three adults and one child.*

In close social situations the participants know their roles. Clark (1996) claimed that the interlocutors' rights, duties, and potential joint purposes are usually quite clear. All they need to establish is the joint purpose for that occasion that they can do with a routine procedure. The first interlocutor initiates the conversational routine often with a phrasal unit, and the second interlocutor completes it by complying. Use of conversational routines and formulas requires shared background knowledge of which there is very little in *lingua franca* communication. Therefore it is quite clear why *lingua franca* communicators avoid formulaic language. For them literality plays a powerful role. But does this really mean that the *idiom principle* works differently when bilinguals use their L1 or L2?

5. Dataset

I conducted another study to examine what happens to speech production of bilingual speakers when they participate in *lingua franca* communication. Will the *idiom principle* be really blocked for them? Or will their language production still be driven by the salience of the *idiom principle* resulting in significant attempts to use formulas rather than freely generated expressions?

I examined the language production of bilingual non-native speakers of English in seven conversations. These conversations were 30-minute recordings of spontaneous speech on topics like health, sports, living in Albany and the like. The participants were as follows: C1 Japanese and Korean, C2 Korean and Turkish, C3 Korean and Chinese, C4 Japanese and Chinese, C5 Chinese and Korean, C6 Korean and Burmese, C7 African-French and Korean. As it can be seen, the participants were mainly Asian speakers with two exceptions. There is a major difference between the research on *lingua franca* described in the previous chapter and this one. In the former project I focused on the general use of all types of formulaic expressions in the conversations. In this research my main focus was on the *idiom principle* and the way bilingual speakers structured certain sequences within the conversations. Two types of production sequences were selected within the 30-minute sessions: “A,” how do participants introduce themselves (closed social situation) and “B” one new topic introduction from each conversation (open social situation) that was usually the first attempt to change the topic in the conversation. The excerpts I used for analysis can be found in the Appendix.

6. Saliency of Formulaic Expressions

How do participants introduce themselves?

Out of the seven conversations, we have introduction in four cases (C1, C3, C4, C7). In the other three cases speakers started in *medias res*, right in the middle. Introduction that is supposed to be a closed social situation requires formulaicity in most languages. Our four examples demonstrate that the *idiom principle* usually works in the L2 if the bilingual speakers are in a well-known closed social situation that exists across cultures. The speakers relied mainly on well-known situation-bound utterances rather than freely generated expressions. For instance:

- *Let me introduce myself first.*
- *So glad to meet you. Let me ask you how long you have been here?*
- *Can I ask your name?*
- *Nice to meet you.*
- *How long have you been here?*

In the other three conversations there is no direct introduction because the subjects knew each other. But the start of conversation in each case shows an endeavor to use formulaic expressions such as

Do you like sports?
What kind of sports?
... do you think there are many activities in Albany...
.... do you keep yourself healthy?

It is important to note that our subjects were students with pre-advanced level of English. They were all familiar with the formulas that are used in introduction in the target language.

Introducing a new topic

New topic introduction is an open social situation. Although the frame is well known, language that is associated with it is much less formalized than in close social situations. As the examples below demonstrate, each subject used mainly some *ad hoc* generated way to introduce the topic. However, the *idiom principle* was still on because the *ad hoc* generated utterance chunks are combined with some formulaic expressions that are relevant to the matter the participants attempted to talk about like in C3B, C6B and C7B.

- C1B:** – Ok it's been three or... three months so far right? Do you like living in Albany? Living in America?
- C2B:** – And my country ... in my country peoples don't like sport.
- C3B:** – So what about here?
- C4B:** – So can you please tell me the difficulties in life here.
- C5B:** – Another thing I noticed about American food is that ... although its contains a lot of fat or something unhealthy, but there's always options you can choose like low ... low calorie grocery or zero calorie version of diet thing
- C6B:** – And what about ... do you care more about ... food? Since this is like another other foreign country ... so do you take care more about choosing some food?
- C7B:** – You say you live in Albany so how is the place where you live? Can you describe the place where you live?

Comparison of the bilingual project with the lingua franca project

It is interesting to compare the numbers in the bilingual project (BP) with those of the *lingua franca* project (LFP). The similarities are striking although the subjects were totally different, their language proficiency was also different, and they talked about different topics. In the BP the subjects worked in pairs, while in the LFP there were 6-7 subjects in each group. In the bilingual project only five groups of formulas were considered, while in the LFP grammatical formulas were also counted.

Table 3. Number of Formulas that Represent the Five Types of Formulas in the Bilingual Project

Fixed Semantic Units	Phrasal Verbs	Speech Formulas	Situation-bound Utterances	Idioms	Total
276	133	227	240	156	1032

Note. Total number of words: 13513; Total number of formulas: 1032; Percentage: 7.63%

Table 4. Number of Expressions that Represent the Six Types of Formulaic Units in the Lingua Franca Study (same as Table 2.)

Grammar Units	Fixed Semantic Units	Phrasal Verbs	Speech Formulas	Situation-bound Utterances	Idioms	Total
102	235	281	250	57	115	1040

Note. Total number of words: 13726; Total number of formulas: 1040; Percentage: 7.57%

7. Discussion and Conclusion

The research revealed some important features of bilingual language use. Based on the presented results we can claim that the *idiom principle* does not seem to depend on how many languages an individual can speak and on what level. The important thing is that *the economy principle affects the use of any language of bilinguals and multilinguals*, the question only is to what extent. Human beings want to achieve as much as possible with the

least possible effort, both in production and comprehension. The best way to do that is to use as many prefabricated chunks of language and possible and combine them with *ad hoc* generated utterances in a creative way. So the hypothesis about the *idiom principle* being blocked in subsequent languages was not supported by this study. However, there is another side of the matter: How can bilinguals cope with the requirements of the *idiom principle* in their L2? Does the principle operate to full extent as in L1? The results of both studies show that not exactly. There are several factors that are not present in L1 but are there in L2 affecting the functioning of the *idiom principle* in different degree. Such factors include language proficiency, willingness to use certain formulas, language fluency of other participants, lack of core common ground, and others. As a result, the actual production of formulaic expressions in the L2 of a bilingual will always be lower than in L1.

As mentioned above, there are several variables in which the two studies differ. However, they show striking similarities in the use of formulaic language in general. The total number of words in both projects is very close: LFP: 13726, BP: 13513, and so are the total number of formulaic expressions: LFP: 1040, BP: 1032. It just cannot be by chance that these numbers are so close. The number of fixed semantic units and speech formulas is also very similar. This refers to the fact that the conclusion of the LFP (Kecskes 2007) was correct when it emphasized that bilinguals in their L2 when participating in *lingua franca* communication prefer the use of semantically more transparent language to formulaic language, so as to make sure that they will be understood by all interactants.

Although the general use of formulaic language is very similar in the two studies, there are still differences in the distribution of formulas under the influence of specific variables mentioned above. For instance, there are differences in the use of phrasal verbs and situation-bound utterances. In the bilingual project the subjects used much more situation-bound utterances than in the LFP. However, the use of phrasal expressions shows a different picture. The use of situation-bound utterances is a sign of the L2 language socialization process (cf. Kecskes 2003; Ortactepe 2012). The bilingual participants had pre-advanced proficiency in English and overall spent more time in the target language environment than the subjects in the LFP whose proficiency level was intermediate. Besides, the subjects in the BP were students while in the LFP participants came from the community.

There is also explanation for the significant difference in the use of phrasal expressions that was used much more frequently by the LFP subjects. As mentioned above, the participants of the *lingua franca* projects came from the Albany community to improve their English in evening classes conducted by TESOL students. Their syllabus put special emphasis on the use of phrasal verbs in English.

Based on the two studies (LFP and BP) we can conclude that, although the *idiom principle* affects any language production, it results in less formulaic language use in L2 than in L1 of bilinguals. This claim basically concurs with the findings of other studies (cf. Bolander 1989; Pawley & Syder 1983; Warga 2008; Weinert 1995) that also talked about the restricted use of formulaic language in L2.

References

- Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word-combinations. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications*, (pp. 101-122). Oxford: Clarendon Press.
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: Lexical bundles in university teaching and textbooks. *Applied Linguistics* 25, 371-405.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Pearson Education.
- Bolander, M. (1989). Prefabs, patterns and rules in interaction? Formulaic speech in adult learners' L2 Swedish. In K. Hyltenstam & L. K. Obler (Eds.), *Bilingualism across the Lifespan: Aspects of acquisition, maturity, and loss* (pp. 73-86). Cambridge: Cambridge University Press.
- Bolinger, D. (1976). Meaning and memory. *Forum Linguisticum* 1, 1-14.
- Cheng, Winnie, Chris Greaves, John McH. Sinclair & Martin Warren (2009). Uncovering the extent of the phraseological tendency: Towards a systematic analysis of congrams. *Applied Linguistics* 30(2). 236–252.
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Coulmas, F. (Ed.). (1981). *Conversational routine: Explorations in standardized communication situations and prepatterned speech*. The Hague: Mouton.

- Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. J. Doughty & M. H. Long (Eds.), *The Handbook of second language acquisition* (pp. 63-103). Malden, MA: Blackwell.
- Ellis, N. C., Simpson-Vlach, R., & Carson, M. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly* 42, 375-396.
- Fillmore, C. J. (1976). The need for a frame semantics within linguistics. *Statistical Methods in Linguistics* 12, 5-29.
- Fillmore, C. J. (1982). Frame semantics. In The Linguistic Society of Korea (Ed.), *Linguistics in the morning calm* (pp. 111-137). Seoul: Hanshin.
- Gairns, R., & Stuart, R. (1986). *Working with words: A guide to teaching and learning vocabulary*. Cambridge: Cambridge University Press.
- Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics*, 7, 183-206.
- Giora, R. (2003). *On our mind: Salience, context and figurative language*. Oxford: University Press.
- House, J. (2003). Misunderstanding in intercultural university encounters. In *Misunderstanding in Social Life: Discourse approaches to problematic talk*, J. House, G. Kasper, & Ross, S. (Eds.), *Misunderstanding in social Life: Discourse approaches to problematic talk*, (pp. 22-56). London: Longman.
- Howarth, P. (1998). Phraseology and second language proficiency. *Applied Linguistics* 19, 24-44.
- Hymes, D. H. (1962). The ethnography of speaking. In T. Gladwin & W. C. Sturtevant (Eds.), *Anthropology and human behavior* (pp. 13-53). Washington: The Anthropology Society of Washington.
- Kecskes, I. (2000). A cognitive-pragmatic approach to situation-bound utterances. *Journal of Pragmatics* 32, 605-625.
- Kecskes, I. (2003). *Situation-bound utterances in L1 and L2*. Berlin/New York: Mouton de Gruyter.
- Kecskes, I. (2007). Formulaic language in English lingua franca. In I. Kecskes & Horn, L. R. (Eds.), *Explorations in pragmatics: Linguistic, cognitive and intercultural aspects* (pp. 191-219). Berlin/New York: Mouton de Gruyter.
- Kecskes, I. (2012). Is there anyone out there who really is interested in the speaker? *Language and Dialogue* 2, 283-297.
- Kecskes, I. (2013). *Intercultural pragmatics*. Oxford: Oxford University Press.

- McEnery, T., & Wilson, A. (1996). *Corpus linguistics*. Edinburgh: Edinburgh University Press.
- Miller, J., & Weinert, R. (1998). *Spontaneous spoken language: Syntax and discourse*. Oxford: Clarendon Press.
- Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical phrases and language teaching*. New York: Oxford University Press. Octopus. (1995, October). p. 144.
- Ortaçtepe, D. (2012). *The development of conceptual socialization in international students: A language socialization perspective on conceptual fluency and social identity (advances in pragmatics and discourse analysis)*. Cambridge: Cambridge Scholars Publishing.
- Pawley, A., & Hodgetts S. F. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. *Language and Communication* 5, 191-226.
- 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.
- Prodromou, L. (2008). *English as a lingua franca: A corpus based analysis*. London: Continuum.
- Segalowitz, N. & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition*, 26, 173-199.
- Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics* 31, 487-512
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Oxford: Blackwell.
- Warga, M. (2005). 'Je serais très reconnaissant': Formulaic vs. creatively produced speech in learners' request closings. *Canadian Journal of Applied Linguistics*, 8, 67–94.
- Weinert, R. (1995). The role of formulaic language in second language acquisition: A review. *Applied Linguistics*, 16, 180–205.
- Wray, A. (1999). Formulaic language in learners and native speakers. *Language Teaching* 32, 213-231.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Wray, A. (2005) Idiomaticity in an L2: Linguistic processing as a predictor of success. In B. Briony (Ed.), *IATEFL 2005: Cardiff conference selections* (pp. 53-60). Canterbury: IATEFL.

Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Language and Communication* 20, 1-28.

Appendix: Excerpts

C1: Korean and Japanese

A

- Can I ask your name?
- I'm Tsubasa.
- Tsubasa, ok
- And your?
- Seungjung, I'm from South Korea.

B

- Ok it's been three or... three months so far right? Do you like living in Albany? Living in America?
- Yes I like..
- What makes you like this life? What is your ... like .. What you like about living in Albany?
- I stay here only 4 month in this semester so I have no time. I go to many place... I went to Boston, Washington DC, of course New York City

C2: Korean and Turkish

A

- Sports. Do you like the sports?
- I like
- What kind of sports?
- I like tennis

- Oh ok
- And soccer
- Soccer
- When I was young I played soccer.

B

- And my country ... in my country peoples don't like sport.
- Oh really
- I like but ... they like but they haven't time. I see in Albany too many people like sport. And they run and fitness.
- Yeah
- They fitness. Too many people play tennis. So I think they sport . they they keep yourselves healthy.

C3: Korean and Chinese

A

- How long have you been here?
- Oh like a.... Getting to be ... almost one year
- One year?
- Yeah, almost one year. But it's like ... ten months. ... since ive been here
- Oh it's good.
- Two months to go..

B

- So what about here?
- Well the only experience I got from here was like hanging out with American college students. And they were like ... I think they were really fun to hanging out ... inside the house. Having drinks inside the house not going out... maybe sometimes go out for a drink....

C4: Chinese and Japanese

A

- Let me introduce myself first. I'm a visiting scholar from a Chinese university. Ok. So we are familiar.
- (laughing)
- Because we are neighbors. Actually ... right. So glad to meet you. Let me ask you how long you have been here?
- Two and a half month.

B

- So can you please tell me the difficulties in life here..
- Ah.
- Or challenges.
- Everything difficult.
- Very. Could you please give me some examples?
- Shopping is difficult.
- Yeah ... really... would you please describe it in detail?
- In [shopright?] ... we have to put on ... my good ... belt ... on the belt ...

C5: Chinese and Korean

A

- I surely ... do you think there are many activities in Albany ... are Do you many activities are in Albany to s.... keep healthy ?
- Yeah, actually on campus you know there is a gym.
- A gym ... ah I heard... I heard students can swim.

B

- Another thing I noticed about American food is that ... although its contains a lot of fat or something unhealthy, but there's always options you can choose like low ... low calorie grocery or zero calorie version of diet thing.
- Right right.
- The diet version you can always choose like....

- Right.
- That's what I select ... what I choose when I have to eat American food.

C6: Korean and Burmese

A

- ok good. ... mmmm ... do you keep yourself healthy? So I just want to ask do you exercise on a regular basis? Here in Albany?
- Ah no. but sometime I did it. But sometime not. It's not always.
- Not always.
- Yeah.

B

- And what about ... do you care more about ... food? Since this is like another other foreign country ... so do you take care more about choosing some food?
- Yeah I choose some food.
- Oh like what?
- Like ... You know in my country like a ... we always eat rice and soup...
- Oh right, you are from ...
- Burma like ...

C7: African French and Korean

A

- My name is Patrick.
- My name is Emi.
- Emi. Nice to meet you.
- Me too. Nice to meet you.
- It's not very formal. You can answer a few questions in an informal way. Just be relaxed.

B

- You say you live in Albany so how is the place where you live? Can you describe the place where you live?
- What? Sorry.
- I mean the place where you live. You said you live in Albany right?
- Place? What place? Korea
- Place ... place...
- Ah place... describe in Albany
- Yeah is it a good place? How is it? How are the houses there? Can you walk out straight and get the first ... do you like the place where you live here in Albany?
- I like living in Albany. Because the Albany is the... especially I [word] almost two months... it's quiet ... nice people ... neighbor...
- Yeah you have nice neighbors.
- Yeah and making [word]