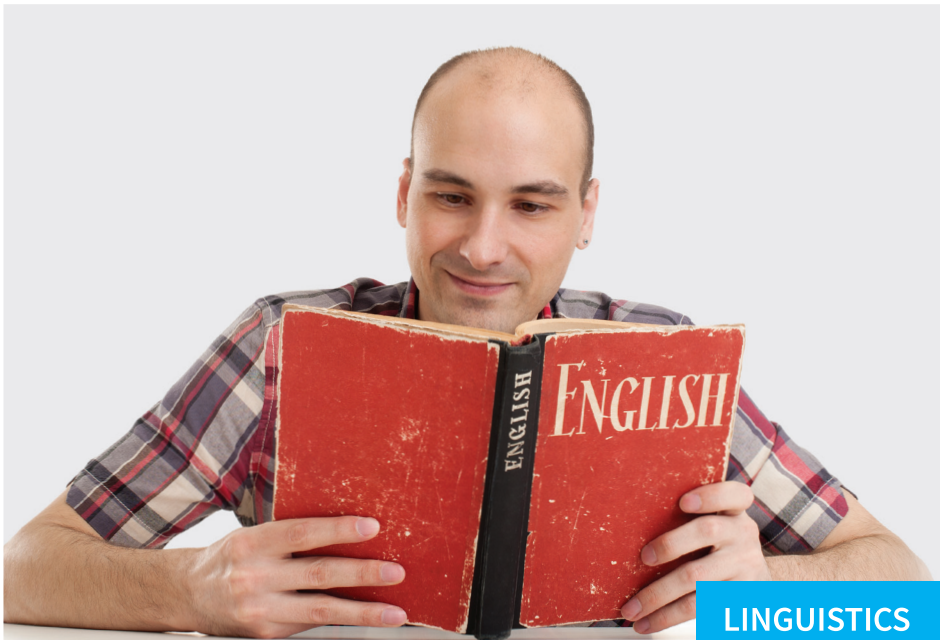


editors

Ewa Waniek-Klimczak
Anna Cichosz

E

Variability in English across time and space



LINGUISTICS

PHONETICS, DIALECTOLOGY, HISTORICAL LINGUISTICS

Formulaic language in native and learner English – a corpus-based study of silent pauses

<http://dx.doi.org/10.18778/8088-065-8.05>

Wojciech Rajtar

University of Lodz

Abstract

The aim of this on-going study is to investigate whether intersegmental silent pauses with regard to formulaic sequences in native speaker English and Polish learner English occur in similar patterns, and whether the selection of the most common two- and three-word phrases in both types of English is alike. The analysis was conducted on the basis of the conversational sub-corpus of the British National Corpus and the spoken part of the PELCRA Learner English Corpus (Pęzik 2012), from which potential lexical bundles were extracted. Then, temporal analysis of audio samples from both corpora corresponding to the potential lexical bundles was performed in order to determine their prosodic features. As a result, the degree of relations between formulaicity of an utterance and its prosodic features, as well as the range of the most frequently used formulaic sequences was shown for both types of English analyzed.

1. Introduction and theoretical background

Fluency is one of the most important aims for learners of languages and language teachers. One of the factors that are believed to improve fluency is the use of formulaic language (van Lancker et al. 1981), sequences of which are “always produced fluently with an unbroken intonation contour and no hesitations for encoding” (Peters 1983: 8). Pawley and Syder (1983) also further elaborate on this assumption stating that formulaic language is used more effectively because

a memorized sequence, even if consisting of several words, is easier to process than a creatively generated utterance of the same length. That is one of the reasons why attention devoted to lexical bundles, “the most frequently occurring sequences in the register” (Biber et al. 2004), and their phonological features has increased recently; although the focus has been mainly directed at child language and EFL learners so far. In these areas, several prosodic features have been proven to be typical of formulaic sequences (e.g. high frequency phrases are more likely to be reduced (Bybee 2002, 2006); formulaic sequences less likely to contain pauses and hesitations (Raupach 1984; Dechert 1983). However, little analysis of such features with respect to adult native English or ESL learner English has been conducted as yet.

1.1. Lexical bundles

The first term that is of importance is what may be considered a unit of formulaic language. Lexical bundles, also called n-grams, were defined by Biber (Biber et al. 2004; Biber et al. 1999) as “the most frequently occurring lexical sequences in a register” and in the last decade they have received a lot of attention from scholars and researchers. Lexical bundles are not always equal with traditional language units like noun phrases or adjectival phrases, and may cross over several structures e.g. *In this study we, should be noted that* (Allen 2009).

One of the features of language that benefits from a speaker’s knowledge and usage of formulaic language is naturalness, which is very important for language learners. What is more, as shown in a study by Neil Millar, misuse of formulaic language can lead to potential communication difficulties (Millar 2009). In the study, Millar measured speaker reading times of collocations taken from native speaker academic writing and Japanese learners’ academic writing. The study showed that those learner collocations that were different from native speaker norms take more time to process in reading.

Lexical bundles are also a convenient research subject because of the method for their discovery. They are extracted purely on the basis of frequency in language through corpus data; in the case of this study – the two spoken language corpora. Because of this method, they can be considered solid, empirical data.

1.2. Native-like selection

The notion of lexical bundles is connected to a large extent to two linguistic capacities discussed in a 1983 essay by Pawley and Syder entitled *Two puzzles for linguistic theory: nativelike selection and nativelike fluency*. The first notion is introduced by the authors as:

the ability of the native speaker routinely to convey his meaning by an expression that is not only grammatical but also nativelike; what is puzzling about this is how he selects a sentence that is natural and idiomatic from among the range of grammatically correct paraphrases, many of which are non-native-like or highly marked usages (191).

The second notion is introduced as:

the native speaker's ability to produce fluent stretches of spontaneous connected discourse; there is a puzzle here in that human capacities for encoding novel speech in advance or while speaking appear to be severely limited, yet speakers commonly produce fluent multi-clause utterances which exceed these limits (191).

Pawley and Syder begin explaining the 'puzzle of nativelike selection' by choosing generative grammar as a reference point for their theory. This stance on grammar proposed by Noam Chomsky (1957) is based on a claim that acquiring a set of rules enables a speaker of a language to produce an infinite number of correct sentences and to distinguish them from ungrammatical ones. This approach focuses on a set syntactic rules called grammar that allow speakers to have a creative power of composing an infinitely large set of grammatical sentences. The authors then argue that if a speaker were to fully exploit the "creative potential of syntactic rules", their utterances would not be considered as showing native-like properties. In reality, most of the possible sentences in a language, even though grammatically correct, when looked at, would be "judged to be 'unidiomatic', 'odd' or 'foreignisms'", while only a small per cent of the infinite set would be accepted as native-like by other speakers of the given language.

To illustrate this assumption, Pawley and Syder (1983: 194) provide a piece of narrative, spoken by an elderly New Zealander remembering his family's circum-

stances at the time World War I broke out, and juxtapose it with a paraphrase of the same narrative that is fully grammatical but highly unnatural:

(1) I had four uncles

they all volunteered to go away

and ah that was one Christmas

th't I'll always remember ...

(2) The brothers of my parents were four

Their offering to soldier in lands elsewhere in the army of our

country had occurred.

There is not a time when my remembering that Christmas will not take place...

This juxtaposition shows that a set of syntactic rules of generative grammar is not all that a language learner has to acquire in order for his speech to be seen as native-like. Apart from that, a language learner needs to “learn a means for knowing which of the well-formed sentences are nativelylike – a way of distinguishing those usages that are normal or unmarked from those that are unnatural or highly marked” (Pawley and Syder 1983: 194).

Making this distinction is particularly difficult for language learners who have learned the grammar of the language from a grammar book without being exposed to the living language in the way it is used by native speakers. Such learners usually produce utterances that, even though grammatically correct, will be perceived as awkward and unidiomatic. The situation is different when one is learning a foreign language by immersion, where a learner acquires a language in a native-like manner. In the process of learning, such a person will be able to acquire the idiomatic use of language along with the command of grammar. The distinction between what is unmarked and marked, or natural and unnatural is what the authors call “the puzzle of nativelylike selection.”

1.3. Native-like fluency

The second puzzle from the essay by Pawley and Syder is called “nativelike fluency”, which is the native speaker’s ability to produce continuous, fluent utterances lasting for ten or even twenty seconds (Pawley and Syder 1983). Such a skill is very difficult to achieve for a language learner, and the process of acquiring it takes a very long time. This may even become a problem for native speakers when they, for example, try to “express (...) thoughts on an unfamiliar subject, or to deliver an unrehearsed monologue to a silent audience, as when tape-recording a letter or radio talk, or when called upon to speak in a public address or formal interview” (Pawley and Syder 198: 199–200). In such cases, brain activity increases, making the speaker ‘look for words’, which often results in phonological and prosodic changes in their speech, e.g. increased number of pauses, more hesitations, changes in tempo.

1.4. Idiom principle vs. open choice principle

John McHardy Sinclair (1991) proposes two principles which determine our strategies in the process of selecting the vocabulary we use in speech. Both principles have much common ground with the notions of native-like selection and native-like fluency. The first one is the idiom principle, which enables a speaker to choose from an array of prefabricated multi-word units. Sinclair (1991: 110) defines this notion so “The principle of idiom is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments.”

The open choice principle corresponds more to the traditional approach of generative grammar, where the speaker combines single lexical items using syntactic rules. Sinclair states that the two principles work alternatively in our speech, and so we produce alternating sequences of spontaneous creative utterances and prefabricated pieces.

1.5. Holistic storage

The two “linguistic puzzles” proposed by Pawley and Syder (1983) show that in order to achieve native-like performance in a foreign language, we need more than just the Chomskyan concept of generative grammar or just the open choice principle. In order to answer the question of what more is needed, they look at the concepts of what they call ‘memorized sequences’, which are sequences of words

that are retrieved from the memory as wholes, as opposed to being composed from individual words in the process of speech. This approach is further supported by phenomena like fossilized errors (Myles et al. 1999), multiword utterances in speech of children (Peters 1983), and fluent chunks (Dechert 1983; Raupach 1984) in the speech of language learners.

At first it might seem that a situation in which speech is just a process of recalling prefabricated data limits the creativity of a speaker that is theoretically enabled by the concept of generative grammar. But it is not necessarily so. As Pawley and Syder (1983: 208) put it:

Possession of a large stock of memorized sentences and phrases simplifies his task in the following way. Coming ready-made, the memorized sequences need little encoding work. Freed from the task of composing such sequences word-by-word, so to speak, the speaker can channel his energies into other activities.

Bolinger (1976: 1) wrote that human language “does not expect us to build everything starting with lumber, nails, and blueprint, but provides us with an incredibly large number of prefabs, which have the magical property of persisting even when we knock some of them apart and put them together in unpredictable ways.” This metaphor accurately demonstrates how formulaicity in a language can help its speakers achieve more efficiency than generative grammar allows for, and also how formulaic sequences, even though consisting of individual words, are closely connected in our mental lexicon.

As far as native speaker language is concerned, a number of factors seem to support the existence of such a holistic manner of storage: semantic non-compositionality (i.e. the meaning of a phrase is not the sum of meanings of the words in it), idiomatic expressions (e.g. *kick the bucket*, *black and blue*), cranberry collocations (e.g. *to and fro*), structures probably passed down from archaic English (e.g. *here comes...*, *believe you me*) (Moon 1998; Lin 2010). According to Lin (2010: 179) the theory of holistic storage

[...] is also in line with the assumption that formulaic sequences should form single intonation units, have less internal dysfluencies such as hesitations and pauses, be uttered faster than rule-based language, and require specific accentual patterns or focus distinction.

Another phenomenon that speaks in favour of the theory of holistic storage are the previously mentioned multiword utterances, or formulaic sequences, in children speech and their phonological properties investigated by Peters (1977, 1983) and Plunkett (1990). A sequence of words is very probable to have been memorized by a child as a whole and stored as a single unit in the mental lexicon when the utterance is characterized by a distinctive intonation pattern (Peters 1977), an unbroken intonation contour (Peters 1983), and articulatory fluency (Plunkett 1990). These factors are criteria that have to be met for a string of words to manifest ‘phonological coherence’, a term which we will discuss later in this paper.

1.6. The frequency-based approach

Another approach used in academic research, which connects formulaic sequences to phonology and prosody is the frequency-based approach. It is concerned with psycholinguistic reasons that influence phonological features of formulaic language. As Bybee (2002) explains it, speech production is a series of “neuromotor routines” that gain efficiency as they are practiced, resulting in a gradually more fluent articulation with each repetition. That is why lexical bundles demonstrate fewer dysfluencies (like pauses or hesitations) and are expected to have increased speech rate and more reductions within their boundaries. In fact, this assumption finds support from Dechert’s (1983) and Raupach’s (1984) observations that second language learners can utter distinctively smooth and fluent stretches that seem to be formulaic amid their other dysfluent productions. As we can see, the theory of holistic storage and the frequency-based approach can work together well, as both look at phonology of formulaic sequences from different angles. That is why Lin (2010) suggests that the two points of view should be combined when investigating this kind of phenomena. The frequency-based approach also indirectly speaks in favour of using corpus data for investigating phenomena related to formulaic language and fluency.

1.7. Phonological coherence

The previously mentioned term of phonological coherence is a notion first used by Peters (1983) as a criterion used in identifying formulaic sequences in the speech of children. According to her, for an utterance to be phonologically coherent it has to fulfil two conditions (the dual criteria): there has to be an unbroken intonation

contour stretching between the boundaries of the sequence, and no hesitations can occur within the sequence. An utterance can be considered phonologically coherent only if the dual criteria are both met. As for formulaic language of adults, a similar assumption is made by Alison Wray (2004), who claims that within a formulaic sequence there should be fewer pauses and errors than in between them, and that lexical bundles are resistant to dysfluency and inaccuracy within them.

The theory of phonological coherence can be considered to be in agreement with the concept of holistic storage. As Wray suggests, such pieces of language (formulaic sequences) are recalled from memory as holistic units. The argument used by her to support this claim is that retrieval of prefabricated formulaic sequences is quicker than recalling a phrase word at a time. That is why retrieving sequences holistically should result in less hesitations, less pauses and a single intonation contour.

In fact, phonological coherence has been used as proof for psycholinguistic processes since as early as 1970s. During this period, studies of prosodic disambiguation were conducted in order to investigate the notions of deep structure and surface structure. A study by Lehiste (1973) aimed at examining the prosodic features that would facilitate disambiguation. In the study she uses sentences like *The old men and women stayed at home* or *The cop killed the robber with a gun* to illustrate how prosodic features can change the deep structure of an utterance. In the first sentence, the deep structure depends on whether a pause is present after the word *men* or not (the men as well as women were old vs. there were old men and some women). In the latter sentence, a pause after the word *robber* is expected if the cop is the one who has the gun. If there is no pause in this place, the robber is the one who has the gun. These examples illustrate how prosodic breaks within utterances delimit phraseological units, which can be considered evidence for a process called chunking.

The psycholinguistic theory behind this process says that chunks are units of language processing (Peters 1983). Lengthy utterances cannot be processed all at once because of the limitations in the 'computing power' of our brains, and that is why such stretches of speech have to be divided into smaller units, chunks, which are just small enough for our short-term memory to process at once. This further gives ground for the claim that the way in which the language is stored should be reflected in phonological and prosodic features of speech.

As Lin (2010) points out, there is a large interdependence between the theories of holistic storage and phonological coherence, and the two notions seem to prove

each other in research. In some research projects, phonological coherence takes the role of the basis for the claim that language is stored holistically (e.g. Wray 2004). In other cases the assumptions are given the other way round (e.g. Moon 1997). It is hard to choose which of the approaches has advantage over the other, but as Lin wrote,

It is perhaps useful to point out that phonological coherence is more of a fact because it is measurable. Holistic storage, however, is more of a claim because its existence is inferred based on facts. On this foundation, we need to critically evaluate on a case to case basis if the evidence is strong enough for us to infer holistic storage from phonological evidence.

More evidence for mutual dependence of formulaicity and phonological features of speech is provided by a 1981 study by Van Lancker, Canter and Terbeek. In the study, native speakers were given a task of reading stretches of text containing specific phrases (e.g. *skating on thin ice*) which carried literal or idiomatic meaning. The researchers found that idiomatic expressions are spoken faster than phrases set in a literal context. What is more, a lower number of pauses occurred in the idiomatic phrases.

This research project, too, complies with the theories of phonological coherence and holistic storage/chunking. The phrases that are idiomatic are articulated faster and uninterrupted because they are retrieved holistically from the mental lexicon, whereas the literal strings are uttered more slowly because they have to be built compositionally, for they are not prefabricated.

In a study of a German learner of English by Dechert (1983), the author arrives at conclusions that the learner's spoken English was riddled with pauses, fillers and hesitations. But what is interesting, he also observed that there were also completely fluent stretches in the recordings of the subject that he called 'islands of reliability' (Dechert 1983: 184). The 'islands' were fragments of speech that the speaker was confident about, probably retrieved holistically, thus causing increases in fluency.

2. Methodology and analysis

The analysis was conducted on the basis of the conversational sub-corpus of the British National Corpus and the spoken part of the PELCRA Learner English Corpus (PLEC) (Pęzik 2012).

The PLEC corpus consists of a written part (2.8 million words), which contains samples of learner English, and a 200,000 word time-aligned spoken subcorpus of learner English, which is the part used for this study. The spoken subcorpus is a series of interviews with learners of English at different education and proficiency levels conducted by academic teachers and PhD students from the English Philology department, University of Łódź.

Data from the British National Corpus was extracted through BNCweb, an online interface for the BNC XML edition written at Lancaster university. The PLEC corpus was accessed through the website of the project (www.pelcra.pl/plec), using the search engine for time-aligned spoken data.

Firstly, the most frequent n-grams were extracted from the PLEC corpus using the formulaic expression browser available at the website of the project:

Table 1. Formulaic expression browser extraction

#	N-gram	Frequency (A)	Independence (B)	B/A	Joint prob	Dispersion	FScore
1	uh_huh	572	59	0.10314685314685315	7.70975	2	52.877613205595935
3	i_don_t_know	310	42	0.13548387096774195	11.5108	29	52.38747509019103
5	for_instance	130	87	0.6692307692307692	8.32041	1	46.05645672539412
6	a_lot_of	178	46	0.25842696629213485	9.38713	30	44.4266802838018
7	for_example	207	51	0.2463768115942029	7.82196	30	43.1284277000129
9	in_poland	125	63	0.504	7.3005	28	39.76749931575577
11	thank_you_very_much	65	30	0.46153846153846156	15.602	9	39.007299814330146
13	i_think_it_s	104	43	0.41346153846153844	9.18749	52	38.74235697577031
15	as_well	77	72	0.935064935064935	7.91003	21	38.41968216096479

Next, 7 most frequent two- and three-word n-grams were selected from the list, with exclusion of *uh huh* for not being a lexical item and also *for instance*, because the phrase was used by just one person (dispersion=1). The same process was applied for the data extracted from BNCweb. Chosen n-grams appear in table 2:

Table 2. N-grams extracted from BNC and PLEC

#	BNC	PLEC
1	I don't know	I don't know
2	isn't it	a lot of
3	that's right	for example
4	or something	in the future
5	you know	in Poland
6	I mean	I think it's
7	come on	as well

The next step of the process was the analysis of audio samples containing the potential lexical bundles seen in the table above. Ten randomly selected (random selection is possible at the BNCweb, in PLEC each third example was chosen) samples of each of the potential lexical bundles were extracted from both corpora, resulting in 140 samples that were subjected to further analysis. The samples analyzed were recordings of language learners only; samples of speech of the interviewees were disregarded.

The data was imported into PRAAT software (Boersma and Weeknink 2014), where the n-grams were isolated from the sound samples. Then, the number of silent pauses was calculated for audio sample and checked against the maximum possible number of intersegmental pauses in the phrase; one possible pause in each two-word phrase, and two possible pauses in each three-word phrase.

3. Results

In the case of the learner English lexical bundles sound samples extracted from the PLEC corpus, there were 110 potential spots where intersegmental silent pauses within the lexical bundles could possibly occur. For the recordings extracted from the BNC spoken subcorpus, there were 80 possible spots for the occurrence of pauses. In PLEC recordings, pauses occurred in 28 of potential spots (24% of possible occurrences). In BNC recordings, pauses occurred in 11 of potential spots (14% of possible occurrences).

The p-value for this data was computed through a Fisher test using R software for statistical analysis, giving a result of $p=0.06806$, showing no statistical significance.

The pauses within the lexical bundles were fairly evenly distributed throughout all the samples analyzed, with the exception of the phrase *I don't know* in both corpora. In the BNC samples most of the pauses occurred in the phrase *I don't know*, whereas in the PLEC samples, the phrase contained no silent pauses whatsoever.

The duration of silent pauses was also measured. In the native English samples, pause duration mean value was 66ms, while in the learner English lexical bundles it was 37ms. The p-value for the Fisher test regarding the duration of pauses between native and learner English is 0,02409, showing statistical significance.

The ranks of occurrence of the n-grams from BNC in PLEC is as follows:

I don't know	-	1
isn't it	-	X
that's right	-	7173
or something	-	63
you know	-	62
I mean	-	17
come on	-	X

(1 – most common n-gram in the corpus, X – n-gram not present in the corpus)

4. Conclusions and discussion

Several conclusions can be drawn from the results of this study. Firstly, it is apparent that the differences in distribution of silent pauses in both types of English analyzed are not significant, thus showing that formulaic sequences are likely to behave similarly in both cases, thus complying with what Wray (2004) wrote as far as pause occurrence is concerned, i.e. pauses are unlikely to occur, and if they do, they are fairly short.

What is interesting, the corpora have only one of the potential lexical bundles in common. This might result from methods used for teaching English in Poland as well as a limited set of interview topics used during the process of gathering data for the spoken part of the PLEC corpus. Such phrases as *for example* can be perceived to have become memorized holistically because of the learning method where the teacher elicits such phrases, even though they do not seem to be that frequent in native speaker English. Also, the format of the interviews, i.e. questions about the learners' plans for the future and about their relations with other cultures is most likely the reason for the occurrence of the phrases *in the future* and *in Poland*.

What is more, most of the n-grams present in native English are rare, or in some cases even nonexistent in Polish learner English (with the exception of *I don't know*). The way English is taught in Poland is most likely the factor to blame for this situation. Textbooks and teaching methods used in Polish schools do not take into account formulaicity and frequency in native language with regard to vocabulary selection. Instead, learners acquire different phrases, which become formulaic for Polish learner English. Such phrases like *for example* or *I think it's* definitely do improve the fluency of the learners' utterances, but at the same time they do not bring them any closer to sounding native-like. What could be done to improve this situation is implementation of the most common native English formulaic sequences into methodology programmes for teaching English – first at the level of input (listening), and then at the level of production as well.

As far as the only n-gram common for samples from both corpora is concerned, there are some interesting observations to be made. Interestingly, the phrase *I don't know* contains the most pauses of the native English bundles analyzed, while no pauses occur in this phrase in learner English. The reason for this polar difference seems to be the way in which the two groups of speakers under scrutiny use the phrase. When we look at the context in which the given phrase was used, native speakers tend to put it in an entirely literal context, exercising the full literal meaning of the phrase (e.g. *I don't know anything about music*), while Polish learners seem to use a completely different strategy; they are likely to insert the phrase into the syntactic clause, devoid of literal meaning (e.g. *some-how you have to, I don't know, maintain it*). This strategy of a “clause-breaking” use of the phrase *I don't know* seems to be used in order for the speakers to “buy themselves time” that they can use for processing the rest of the utterance. The same use of the lexical bundles is visible among native speakers as well, but they

choose different phrases in such contexts; e.g. the phrases *you know* or *or something* seem to easily adopt this “clause breaking” function. A possible reason for the presence of this phenomenon in the speech of Polish learners of English is that the Polish counterpart of the phrase – *nie wiem* – seems to very frequently adopt this function in the Polish language. Because of this, Polish learners subconsciously “calque” the function of the memorized sequence present in their L1 onto its English counterpart.

This is a preliminary study, which is a part of a larger on-going project. The study has several limitations which will be considered during further research in this field. The sample for the study will be extended, and selected so that the data for analysis is as objective as possible. Other phonological features of formulaic sequences, such as filled pauses and intonation contour are to be considered in the full study as well.

References

- Allen, D. 2009. Lexical Bundles in Learner Writing: An Analysis of Formulaic Language in the ALESS Learner Corpus. *Komaba Journal of English Education* 1: 105–127.
- Biber, D., S. Conrad, and V. Cortes. 2004. If you look at...: Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics* 25 (3): 371–405.
- Biber, D., S. Johansson, G. Leech, S. Conrad, and E. Finegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow, Essex: Pearson Education Ltd.
- Boersma, P. and D. Weenink. 2014. Praat: doing phonetics by computer [Computer program]. Version 5.4.01, retrieved 9 November 2014 from <http://www.praat.org/>.
- Bolinger, D. 1976. Meaning and memory. *Forum Linguisticum* 1: 1–14.
- Bybee, J. 2002. Phonological evidence for exemplary storage of multiword sequences. *Studies in Second Language Acquisition* 24: 215–21.
- Bybee, J. 2006. From usage to grammar: the mind’s response to repetition. *Language* 82(4): 711–33.
- Chomsky, N. 1957. *Syntactic Structures*. Paris: Mouton.
- Dechert, H.W. 1983. How a story is done in a second language. In *Strategies in interlanguage communication*, eds. C. Faerch, and G. Kasper, 175–95. London: Longman.
- Lehiste, I. 1973. Phonetic disambiguation of syntactic ambiguity. *Glossa* 7: 107–122.
- Lin, P.M.S. 2010. The phonology of formulaic sequences: A review. In *Perspectives on formulaic language: Acquisition and communication*, ed. D. Wood, 174–193. London: Continuum.

- Millar, N. 2009. Assessing the processing demands of learner collocation errors. Poster presented at Corpus Linguistics 2009, Liverpool, UK.
- Moon, R., 1998. *Fixed Expressions and Idioms in English*. Oxford: Clarendon Press.
- Myles, F., R. Mitchell, and J. Hooper. 1999. Interrogative Chunks in French L2: A Basis for Creative Construction? *Studies in Second Language Acquisition* 21: 49–80.
- Pawley, A., Syder, F. 1983. “Two Puzzles for Linguistic Theory: Nativelike Selection and Nativelike Fluency.” *Language and Communication* 191: 225.
- Peters, A.M., 1977. Language learning strategies: does the whole equal the sum of the parts? *Language* 53(3), 560–573
- Peters, A. 1983. *The Units of Language Acquisition*. Cambridge: Cambridge University Press.
- Pęzik, P. 2012. Towards the PELCRA Learner English Corpus. In *Corpus Data across Languages and Disciplines, Lodz Studies in Language. Vol. 28*, ed. P. Pęzik. Frankfurt am Main: Peter Lang. Forthcoming.
- Plunkett, K., and S. Stromqvist. 1990. The Acquisition of Scandinavian Languages. *Gothenburg Papers in Theoretical Linguistics* 59.
- Raupach, M. 1984. Formulae in second language speech production. In *Second language productions*, eds. H.W. Dechert, D. Möhle, and M. Raupach, 114–37. Tübingen: Gunter Narr.
- Sinclair, J. 1991. *Corpus, concordance, collocation: Describing English language*. Oxford: Oxford University Press.
- The British National Corpus, version 3 (BNC XML Edition). 2007. Distributed by Oxford University Computing Services on behalf of the BNC Consortium. Retrieved from <http://www.natcorp.ox.ac.uk/>, 17.11.2014.
- Van Lancker, D., J. Canter, and D. Terbeek. 1981. Disambiguation of ditropic sentences: Acoustic and phonetic cues. *Journal of Speech and Hearing Research*, 24: 330–335.
- Wray, A. 2004. ‘Here’s one I prepared earlier’: formulaic language learning on television. In *The acquisition and use of formulaic sequences*, ed. N. Schmitt, 249–268. Amsterdam: John Benjamins.

LINGUISTICS

PHONETICS, DIALECTOLOGY, HISTORICAL LINGUISTICS

Synchronic variability in the area of phonetics, phonology, vocabulary, morphology and syntax is a natural feature of any language, including English. The existence of competing variants is in itself a fascinating phenomenon, but it is also a prerequisite for diachronic changes. This volume is a collection of studies which investigate variability from a contemporary and historical perspective, in both native and non-native varieties of English. The topics include Middle English spelling variation, lexical differences between Middle English dialects, Late Middle and Early Modern English forms of address, Middle English negation patterns, the English used by Polish immigrants living in London, lexical fixedness in native and non-native English used by Polish learners, and the phenomenon of phonetic imitation in Polish learners of English. The book should be of interest to anyone interested in English linguistics, especially English phonetics and phonology as well as history of English, historical dialectology and pragmatics.

The book is also available
as an e-book



WYDAWNICTWO
UNIwersytetu
ŁÓDZKIEGO

ul. Williama Lindleya 8
90-131 Łódź

tel.: 42 66 55 863
e-mail: ksiegarnia@uni.lodz.pl

ISBN 978-83-8088-065-8



9 788380 880658