Wojciech Witkowski (https://orcid.org/0000–0001–8873–9819)
Institute of English Studies
University of Wrocław

Paulina Witkowska (https://orcid.org/0000–0002–5393–7193)
Institute of Polish Studies
University of Wrocław

Feasibility of using Web-based occurrences in tracing lone items' diffusion

1. Introduction

This paper focuses on presenting insights regarding the diffusion of English-origin verbs in Polish based on the quantitative data obtained from the Internet. The current consensus regarding the hierarchy of lexical borrowing treats verbs as least prone to become loan elements (see Haspelmath 2009 and sources cited therein). This, in turn, makes their occurrence in Recipient Language a valuable source of information regarding the process of borrowing or code-switching.

Loan items' diffusion, i.e. spread in usage among members of a speech community, has been discussed in research focused on borrowing or code-switching quite extensively. Apart from being given a qualitative account, it has also been researched from a quantitative perspective. Poplack, Snakoff, Miller's (1988) seminal work on French-English loans, Treffers-Daller's (1994) work on French and Dutch; van Hout and Muysken's (1994) research on Spanish and Quechua, Kuźniak's (2009) work on English-Polish loans or Polack and Dion's (2012) diachronic research on French-English switches can be recalled here as illustrating the abovementioned qualitative paradigm.

The quantitative approach is adapted likewise, in the paper at hand with the intention of testing if approaches present in the field can be applied to data gathered from Google search results. Using Internet-based data may seem to stand in clear contrast to previous research, which focused on data elicited from community members. However, on closer inspection, each Google search hit, which results from locating a text string matching the term searched for on a web-page, can be treated as an individual (written) utterance created by an individual speaker.

This paper is organized as follows. Section 2 reviews Poplack, Snakoff, Miller's (1988) and Polack, Dion (2012) research, which serves as the conceptual basis for the methodology explored in this paper. Section 3 presents the English-origin verbs used in the current research and the collected Google hits frequencies. Sections 4 and 5 discuss the method used to transform the collected raw frequencies

and the results of applying the method developed in this paper to the collected data. Section 6 presents the potential implications resulting from the current research.

2. Nonce Borrowing Hypothesis and diffusion of other-language items

The research presented in this paper builds conceptually on the approach to investigating the process of language mixing originally laid forward in Poplack, Sankoff, Miller (1988) and adapted in Poplack, Dion (2012), Poplack (2012, 2018) under what was termed *Nonce Borrowing Hypothesis*. This approach allows to study the introduction of a foreign element to a Recipient Language (RL) and its subsequent development in RL quantitatively by focusing on contemporary developments, i.e. looking at language from a synchronic perspective and studying them in their relatively small time span, i.e. applying diachronic perspective on what could be characterized as a micro-scale.

Poplack, Sankoff, Miller (1988) develop *Nonce Borrowing Hypothesis* (NBH) on the study of French-English code-switching in Ottawa-Hull region. They focus on quantitative, qualitative, and social aspects of the presence of English language Lexical Items (LIs) (lone other-language items (LOLIs) in their terminology) in the speech of a francophone community. The results of their research allow them to formulate NBH in the following manner (see also Poplack 2012): (i) RL language users code-switch or borrow spontaneously; (ii) with the exception of phonological adaptation, which can remain variable, Donor Language LIs abruptly adapt to morphological and syntactic requirements of the RL; (iii) morphological and syntactic adaptation largely precedes social acceptance, i.e. Donor Language LIs are well adapted before they show diffusion across a community of speakers. Additionally, on quantitative grounds Poplack, Sankoff, Miller (1988) provide a four-party distinction between types of borrowed elements: (i) idiosyncratic – used only by one speaker, (ii) nonce – occurring only once in the corpus, (iii) recurrent– occurring at least 10 times in the corpus, and (iv) widespread – used by at least 10 speakers (see also Poplack 2012, 2018). Additionally, as noted by Poplack, Sankoff, Miller (1988: 57), by conjunction, idiosyncratic borrowed types are nonce borrowed types, and widespread borrowed types are frequent borrowed types.

Poplack, Dion (2012) adapt the methodology proposed by Poplack, Sankoff, Miller (1988) and apply it to a research-oriented English language LIs present in the speech of Quebec-French speakers born between 1846–1994, recorded in the data ranging from 1946 to 2007. The specific point of interest of Poplack, Dion's research is nonce borrowings defined in terms of Poplack, Sankoff, Miller (1988), i.e. those English language LIs that were uttered exactly once by exactly one speaker. The analysis of their diachronic data is oriented with respect to three assumptions that are presented below:

- **Diffusion assumption** 'Lone other-language items introduced as nonce words typically increase in frequency and diffusion.' (Poplack, Dion, 2012: 281)

- **Graduality assumption** 'Lone other-language items are introduced in donor-language phonological, morphological, and syntactic form (i.e., as code-switches), but in tandem with their increase in frequency and diffusion, they are gradually integrated into the recipient-language structure (become bona fide loanwords)' (Poplack, Dion, 2012: 281)
- **Identity Assumption** 'At any given point in time, and possibly throughout the process, single-word code-switches cannot be distinguished from (nonce) borrowings.' (Poplack, Dion, 2012: 281)

In the course of their analyses, Poplack and Dion reject the Graduality assumption by showing that on the basis of such diagnostics as adapting French verbal morphology and French determiner realization rules, lone other-language items abruptly adapt to the grammatical requirements of the RL. This result is in line with the synchronic research of Poplack, Sankoff, Miller (1988). Next, relying on the diagnostics used for testing the Graduality assumption but applying them to the comparison between code-switches, nonce borrowings and more frequent borrowings, Poplack and Dion uphold the Identity assumption. This is motivated by the fact that no significant differences were discovered between these groups of LIs. Lastly, as for the Diffusion assumption, Poplack and Dion (2012: 286) argue for the following types of nonce borrowings: (i) ephemeral – occur only once and do not recur; (ii) inveterate – recur only once; (iii) advancing – recur and exhibit increasing frequency[1].

Poplack, Sankoff, Miller's (1988) and Poplack, Dion's (2012) research offer an effective approach to analyzing both the process and the products of language mixing that allows the capture of its synchronic and diachronic facets. However, on closer inspection, their quantitative approach, which serves as the basis for the classification of lone other-language items, reveals certain inconsistencies. Firstly, from a synchronic perspective, Poplack, Sankoff, and Miller's (1988) treatment of borrowed types creates the following categories:

- idiosyncratic = used exactly by one speaker regardless of number of attestations
- nonce = used exactly once by exactly one speaker
- widespread = used by at least 10 speakers regardless of number of attestations
- recurrent = attested at least 10 times regardless of the number of speakers

Such a classification naturally creates more than 4 possible groups. For instance, one can think of widespread: nonce types, i.e. nonce types that are used exactly once but by at least 10 different speakers. This, in turn, enforces the presence of a widespread non-nonce type, i.e. lone other-language items that are used more than once by at least 10 different speakers, which would, under Poplack, Sankoff and Miller's (1988) approach, classify as recurrent type. Moreover, recurrent types appear to cover several other types. For example, apart from pure recurrent items and widespread items that by Poplack, Sankoff and Miller's (1988) definitional criteria are recurrent, this category could also include idiosyncratic type that is attested at least 10 times

---

[1] It should be noted that Poplack and Dion (2012) treat any item that recurs at least one other time as advancing.

in a speech of a single speaker or what has been above dubbed as widespread: nonce type. Accordingly, in effect, Poplack, Sankoff, Miller's (1988) classification seems to be more granular than originally presented. This granularity is naturally desired with respect to measuring diffusion; however, without very precise boundaries for marking each category, its application may seem biased. Secondly, from a diachronic perspective, Poplack, Dion's (2012) seems to fail to account for three other potential frequency of occurrence patterns: (i) items that recur more than once, items that show decreasing frequency, (iii) items whose frequency does not allow to unambiguously determine whether the pattern is either increasing or decreasing. Accordingly, it seems to focus on only one of a variety of possible effects of language change.

3. English-origin verbs in Polish: Data

English-origin verbs analyzed in this study were adapted from Kuźniak and Mańczak-Wohlfeld (2014) where they constitute a subset of collection of nonce borrowings (after Chłopek 2011) or meteors (after Kuźniak 2009) collected by Kuźniak as a part of an online project 'Catch a falling star' carried out between 2010 and 2012 at the Institute of English Studies, Wrocław University. The project was aimed at Institute students and competent language speakers with the intention of eliciting words of English origin that they either actively incorporate or have witnessed being incorporated into Polish language spoken or written texts. From the perspective of the current research, focusing on the items that were elicited a decade ago provides an opportunity to conduct an analysis that can capture both the synchronic aspect (a single generation of speakers) and the diachronic aspect (10-to-12-years time span) of language mixing phenomenon.

Initial list of English-origin verbs comprised of 43 items[2]: apdejtować (IMPERF. update), apgrejdować (IMPERF.upgrade), czejsować (IMPERF.chase), debunkować (IMPERF.debunk), dissować (IMPERF.diss), ekspadować (PERF.expand), fajlować (IMPERF.file), fapować (IMPERF.fap), followować (IMPERF.follow), hajpować (IMPERF.hype), hejtować (IMPERF.hate), kancelować (IMPERF.cancel), keszować (IMPERF.cash.out), komitować (IMPERF.commit), linkować (IMPERF.link), luknąć (PERF.look), obrandować (PERF.brand), obrendować (PERF.brand), otagować (PERF.tag), ownować (IMPERF.own), pullować (IMPERF.pull), puszować (IMPERF. push), requestować (IMPERF.request), riskitować (IMPERF.risk.it), saportować (IMPERF.supprot), sczekałtować (PERF.check.out), shipować (IMPERF.ship (a relationship)), skilować (IMPERF.equip.with.a.skill), spojlerować (IMPERF.spoilerN) szitować (IMPERF.shit), tuningować (IMPERF.tuningN), tweetnąć (PERF.tweet), updejtować (IMPERF.update), upgrejdować (IMPERF.upgrade), wylajtować (PERF. lightA), zabookować (PERF.book), zabukować (PERF.book), zafajlować (PERF.file), zakomitować (PERF.commit), zalajkować (PERF.like), zawrapować (PERF.wrap), zfriendzonować (PERF.friendzone), zreszarować (PERF.reshare).

---

[2]   Spelling variants were treated as separate items. PERF.=perfective, IMPERF.=imperfective

For each of the English-origin verbs, a number of hits were obtained using the Google search engine. For 5 verbs (i.e. riskitować (IMPERF.risk.it), szitować (IMPERF.shit), zafajlować (PERF.file), zawrapować (PERF.wrap), zreszarować (PERF.reshare)) no Google hist were returned. Searches were carried out between July 1, 2023 and July 13, 2023. Queries included all word forms that could be expected given conjugation paradigms of verbs in Polish. As evidenced by Kuźniak, Mańczak-Wohlfeld, 2014) and replicated in the data collected for the current research, items showed morphological adaptation in terms of verbal morphology; thus they can be considered to lend additional evidence in favor of assuming abrupt adaptation to RL, which is in line with the results obtained by Poplack, Sankoff, Miller (1988) and Poplack, Dion (2012). Search options were restricted to Polish language sites only. Counts were gathered for each year – from 2010 to 2022 – separately by checking the search results on the last page of the output. This was done to retrieve the most precise number of hits (by default the hit-count generated by Google on the first result page is a very rough estimate[3]. The collected counts of Google hits were, in fact, the counts for the number web-pages on which the English-origin verbs were found, regardless of how many times the given verb was used on that web-page. As such, the collected hits can be treated as illustrating the number of different texts (which is treated as equivalent of speakers in Poplack, Sankoff, Miller's (1988) and Poplack, Dion's (2012) research) in which the verb under investigation is used at least once. This approach allows, in turn, to distinguish between what in Poplack, Sankoff, Miller's (1988), Poplack's (2012, 2018) or Poplack, Dion's (2012) terms would be nonce lone other-language items (exactly one Google search hit) and widespread lone other-language items (at least ten Google search hits) on one hand, and ephemeral (exactly one Google search hit), inveterate (exactly two Google search hits one for each of two different years), and advancing (increasing number of Google search hit) nonce borrowings on the other hand. **Accordingly, by focusing on the number of web pages, the data collected in the current research allows to measure the diffusion of the English-origin verbs in Polish.**

Obtained Google hits for the individual items were characterized by visible disproportions, from as low as below 0 hits to as high as over 180 hits within the years under investigation. Such disproportions are to be expected given that the initial list of items was composed of English-origin items that, under Poplack, Dion's (2012) assumptions, would range from ephemeral nonce borrowings to widespread lone other-language items. Raw frequencies of Google hits for the verbs considered in this research are presented below in Table 1. Where possible, the items were classified into groups distinguished on the basis of Poplack, Sankoff, Miller (1988), Poplack (2012), Poplack, Dion (2012) and Poplack (2018) criteria.

Table 1: English-origin verbs: Google hits frequency

---

[3] See: https://developers.google.com/search/docs/fundamentals/how-search-works https://support.google.com/websearch/thread/128011128/how-can-i-see-my-number-of-search-results?hl=en).

Type	Verb	Frequency per year (raw Google hits 2010 – 2022)												
		2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022
Ephemeral, Nonce, Idiosyncratic	zakomitować	0	0	0	0	0	0	1	0	0	0	0	0	0
Ephemeral	failować	0	0	0	0	0	0	2	0	0	0	0	0	0
	sczekałtować	0	0	0	0	0	0	0	0	0	2	0	0	0
Inveterate	ekspadować	0	0	0	0	1	0	0	2	0	0	0	0	0
Advancing	apdejtować	2	3	2	4	4	10	6	10	7	7	8	32	9
	apgrejdować	3	3	6	10	3	7	5	10	10	10	8	9	10
	debunkować	0	1	0	0	0	3	5	5	3	4	7	10	15
	kancelować	0	0	0	1	0	0	3	2	0	0	0	4	9
	obrendować	0	0	0	0	4	4	3	3	3	5	10	6	7
	puszować	2	2	4	3	7	6	6	7	8	5	8	7	10
	requestować	0	0	0	0	0	0	0	4	1	4	5	10	7
	updejtować	3	2	1	1	1	4	7	4	4	10	10	5	10
	zabookować	3	2	0	0	2	2	2	4	6	4	4	10	9
	zabukować	2	1	2	4	5	3	2	9	10	6	7	7	10
Widespread, Advancing	dissować	19	20	21	29	34	39	35	40	60	66	71	62	66
	followować	3	5	6	14	23	39	35	43	60	79	93	98	97
	hajpować	3	0	3	5	7	10	10	5	12	15	19	19	26
	hejtować	53	57	75	104	98	139	143	144	142	182	188	170	165
	linkować	93	124	118	123	136	135	143	142	164	168	175	167	169
	obrandować	1	3	4	9	10	18	23	25	26	42	60	61	75
	otagować	41	51	98	89	120	116	129	130	124	135	136	156	131
	tuningować	7	8	6	14	10	3	6	7	9	9	9	22	20
	tweetnąć	38	36	44	44	58	54	61	78	69	59	62	87	71
	upgrejdować	4	0	3	5	10	6	9	4	14	10	17	16	14
Other – not falling under above classifications	czejsować	0	0	0	1	3	2	1	0	2	0	3	2	1
	fapować	6	9	15	17	20	24	18	19	19	26	24	20	16
	keszować	6	6	9	13	14	18	16	19	21	14	20	12	15
	komitować	0	0	0	0	2	1	1	3	5	6	10	3	0
	lunkąć	5	2	0	8	5	3	4	2	5	3	2	4	3
	ownować	3	2	6	0	0	0	3	3	1	2	4	5	6
	pullować	3	0	3	2	4	2	4	6	2	4	6	8	7
	saportować	0	1	1	0	6	3	1	2	4	5	2	4	7
	shipować	4	3	6	10	13	82	57	70	49	37	30	51	60
	skilować	6	6	5	7	7	6	8	13	8	6	10	9	6
	spojlerować	0	0	0	0	3	9	10	6	9	10	5	9	10
	wylajtować	3	0	0	2	2	1	1	0	1	1	1	0	0
	zalajkować	0	2	1	5	3	3	5	5	7	4	4	3	2
	zfriendzonować	0	0	0	0	1	0	0	0	0	1	0	0	1

As can be seen in Table 1, the criteria for distinguishing different types of lone other-language items proposed by Poplack, Sankoff, Miller (1988) and Polack, Dion (2012) allow only partially to unambiguously classify the English-origin verbs used

in the current study. Firstly, with respect to the group labelled Widespread in Table1, it has to be noticed that even under a very liberal approach adapted here, i.e. one that assumes that an English-origin verb is widespread if it occurs more than 10 times at least once, the English-origin verbs listed as Widespread in Table 1 diverge from this pattern in a number of years under investigation. Secondly, and more importantly, a rather large number of tested English-origin verbs that occur in Polish were classified in Table 1 as Other. These items fail to conform to any of the categories proposed by Poplack, Sankoff, Miller (1988) and Polack, Dion (2012) due to the following reasons:

(i) recur more than once e.g. czejsować (IMPERF.chase), wylajtować (PERF. lightADJ), zfriendzonować (PERFfriendzone);

(ii) they show characteristics of advancing nonce borrowings or widespread lone other-language items in the initial part of the time span but drop down in their frequency in later part of it, e.g. fapować (IMPERF.fap), keszować (IMPERF.cash), skilować (IMPERF.equip.with.a.skill), zalajkować (PERF.like);

(iii) they exhibit fluctuation in their frequency of occurrence disallowing for determining its pattern as either advancing or otherwise, e.g. luknąć (PERF.look), ownować (IMPERF.own), pullować (IMPERF.own), saportować (IMPERF.support).

4. English-origin verbs in Polish: Method

As was presented above, relying on raw frequencies of occurrence and utilizing already existing classifications of other-language items provides only a partial account for the distribution of the English-origin verbs scrutinized in this research. These results are primarily based on the nature of the collected data, not on the design of the classifications reviewed above. As is apparent from the preceding sections, this research deviates largely from Poplack, Sankoff, Miller (1998) or Poplack, Dion (2012) in terms of the data sources, and especially in terms of being able to account for the frequency of lone other-language item by a given speaker. Additionally, as was mentioned in Section 3, this study makes a general working assumption that the generated Google hit corresponds to a single use of a lone other-language item by a single speaker. Naturally, this is usually not the case given the websites, i.e. a given web-site may exemplify a number of usages of a given lone other-language item by a number of speakers – a number that will not show up in the count generated by Google hits.

Therefore, as the current study constitutes a rather robust approach to other-language item's diffusion in RL; robust in the sense that a single count is equated with a single use by a single speaker, it requires a modified method of measuring diffusion. The method of measuring diffusion that this research adapts relies on standardizing the observed frequencies. By converting the observed frequencies to standardized points, the applied method lessens the effect of the measurement error indicated above, i.e. the fact that a single count can actually represent more than one use of an English-origin verb by a distinct speaker in such a manner that if the actual counts increase in their values

that increase will be reflected in the increase of the value of the sample's positional measure that serves as the basis for score standardization. Accordingly, to further avoid potential unwanted measurement error effects, on the one hand, and to avoid the influence of high-frequency items found in the sample on the values of standardized scores, on the other hand, the observed frequencies were transformed to robust z-scores, i.e. their values were based on the sample median and Median Absolute Deviation (MAD) in the sample, using the following formula:

$$1.4826 \frac{observed frequency - median}{MAD} = 1.4826 \frac{observed frequency - median}{median(|(observed frequency - median)|)}$$

It has to be stressed at this point that converting the raw frequencies to robust z-scores does not change the conceptual approach to the observed data, but allows one for an outlook on the data that less dependent on raw observation and the sample itself. As in the previous research cited in this paper, higher diffusion items will still exhibit larger z-scores. However, in contrast to using raw figures, using z-scores provides an opportunity to use commonly accepted threshold values to distinguish between those values that are more common and those that are exceptional (outlying in statistical terms). Accordingly, in the current case, any observations for which the robust z-score lies outside the (-3.5 , 3.5) range is treated as indicating excessively lower / higher values than those for the other observations in the sample. In terms of the research presented in this paper, an English-origin verb characterized, for instance, with robust z-score values greater than 3.5 will be treated as definitely diffused among Polish speakers. Consequently, the approach adopted here manages to fall in line with the diffusion assumption (see Poplack, Dion 2012).

5. English-origin verbs in Polish: Results

Robust z-scores were calculated using R software (R Core Team 2018), taking into account all English-origin verbs presented in Table 1. The scores served as the basis for distinguishing between those verbs that can be thought of as being either (i) **diffused** in Polish (widespread in Poplack, Sankoff, Miller's (1988) terminology); (ii) **post-onset stage** (advancing in Poplack, Dion's (2012) terminology); (iii) **onset stage**; (iv) **withdrawal stage**; or (v) **pre-onset stage** (idiosyncratic, ephemeral and inveterate in Poplack, Sankoff, Miller's (1988) and Poplack, Dion's (2012) terminology) with respect to their diffusion. Assignment to a particular group was based on the distribution of robust z-scores with respect to a z-score of 0[4]. Graphical visualization of z-scores for the tested English-origin verbs between 2010 and 2022 is presented below in Figure 1.

---

[4]   In the sample analyzed in this study, robust z-score of 0 equals to 5 occurrences. ±1 robust z-score equals to ±4 observations. The minimal obtained robust z-score is -1.48 and equals to 0 occurrences.

Figure 1: Robust z-scores for English-origin verbs. (Points mark z-score values, red line marks the z-score of 3.5.)
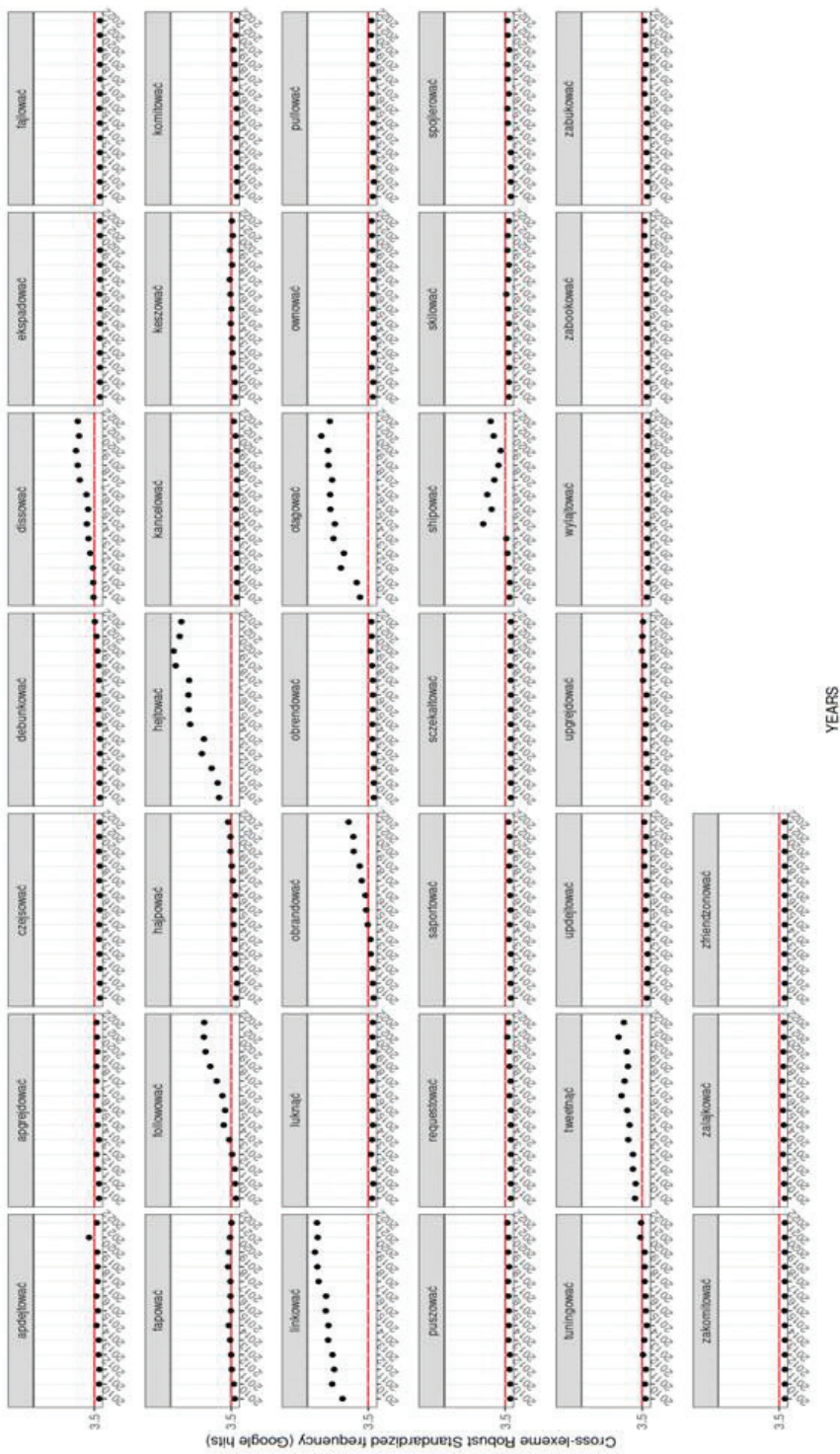
As can be seen in Figure 1, given the threshold of ±3.5 robust z-score, only 9 of tested English-origin verbs can be treated as characterized by an excessively high diffusion. These items are: dissować (IMPERF.diss), fapować (IMPERF.fap), followować (IMPERF.follow), hejtować (IMPERF.hate), linkować (IMPERF.link), obrandować (PERF.brand), otagować (PERF.tag), shipować (IMPERF.ship (a relationship), tweetnąć (PERF.tweet). Accordingly, these English-verbs verbs can be regarded as **diffused.**

The remaining English-origin verbs presented in Figure 1 exhibit robust z-score values below 3.5. Therefore, it becomes necessary to zoom in on their score distribution. Inspection of robust z-score values for English-origin verbs whose z-scores fall below or around 3.5 constitute a heterogeneous group. Based on their robust z-scores distribution, the following sub-groups can be identified:
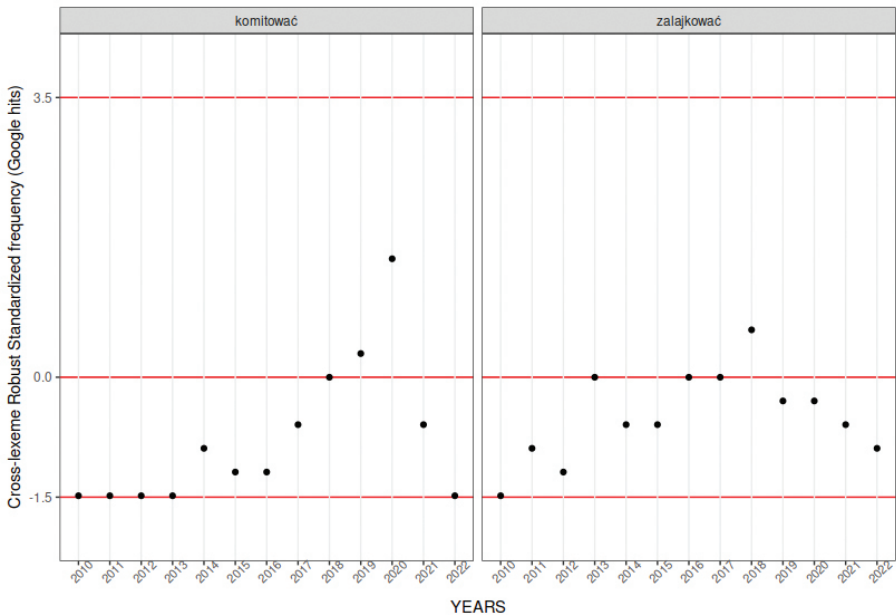
(i) **post-onset stage** English-origin verbs– gradually move to oscillate between the z-scores of 0 and 3.5; possibly reach values greater than 3.5 (see Figure 2): debunkować (IMPERF.debunk), hajpować (IMPERF.hype), tuningować (IMERF. tuningN), upgrejdować (IMPERF.upgrade) (4);
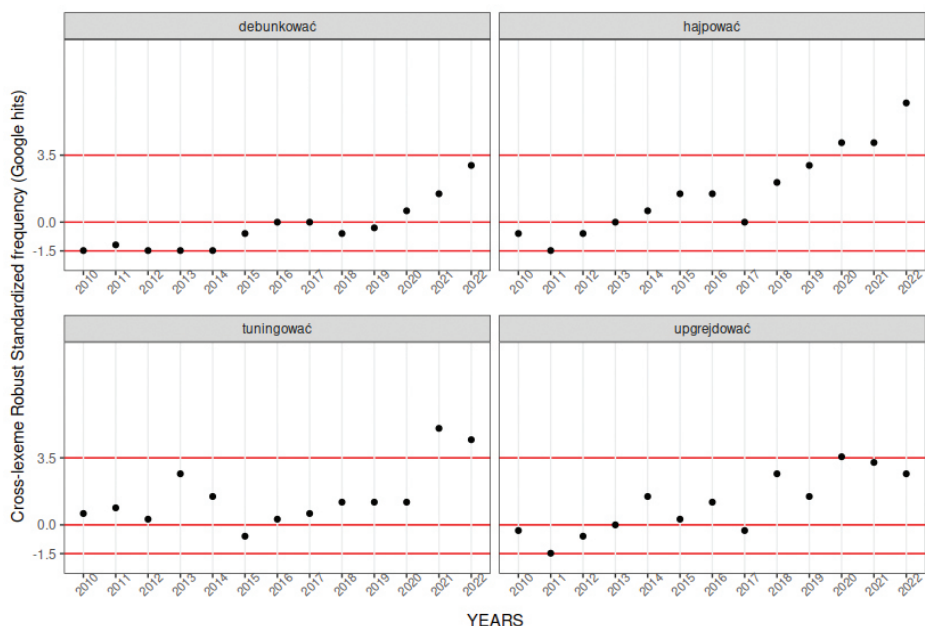

(ii) **onset stage** English-origin verbs – oscillate around the z-score of 0 and do not move towards neither reach z-score of 3.5 (see Figure 3): apdejtować (IMPERF.update), apgrejdować (IMPERF.upgrade), kancelować (IMPERF.cancel), luknąć (PERF. look), obrendować (PERF.brand), ownować (IMPERF.own), pullować (IMPERF.pull), puszować (IMPERF.push), requestować (IMPERF.request), saportować (IMPERF. support), skilować (IMPERF.equip.with.a.skill), spojlerować (IMPERF.spoilerN), updejtować (IMPERF.update), zabookować (PERF.book), zabukować (PERF.book) (15);
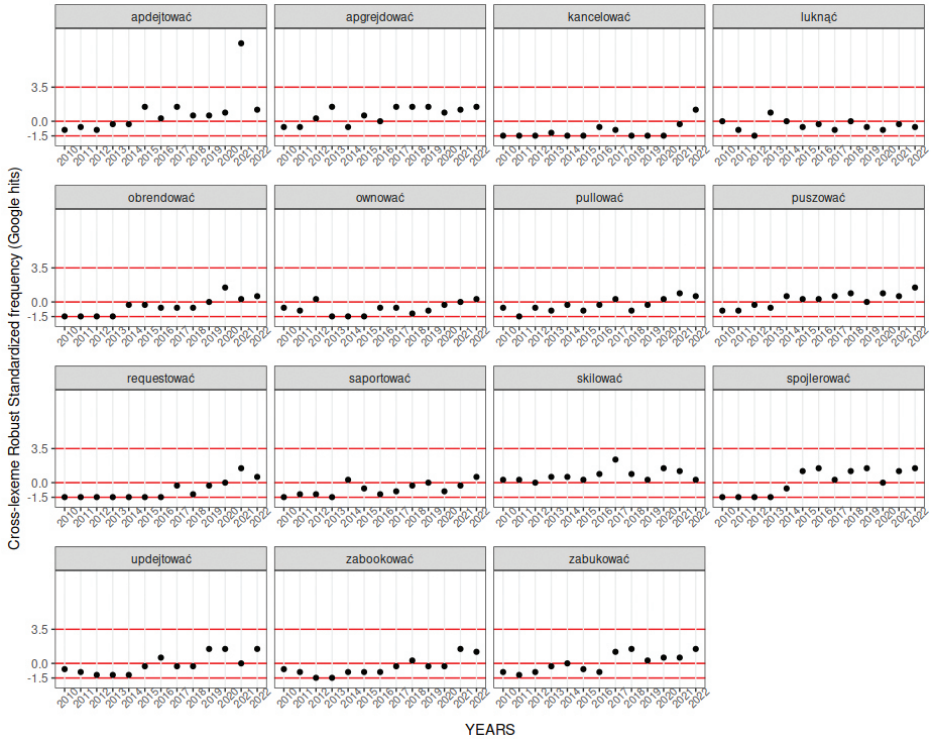
(iii) **withdrawal stage** English-origin verbs – initially oscillate around the z-score of 0 or higher but move towards the minimum value z-score (see Figure 4): komitować (IMPERF.commit), zalajkować (PERF.like) (2);

(iv) **pre-onset stage** English-origin verbs – diverge slightly towards the positive side from the minimum z-score value do not reach 0 z-score (see Figure 5): czejsować (IMPERF.chase), ekspadować (PERF.expand), fajlować (IMPERF.file), sczekałtować (PERF.check.out), wylajtować (PERF.lightADJ), zakomitować (PERF.commit), zfriendzonować (PERF.friendzone) (7).
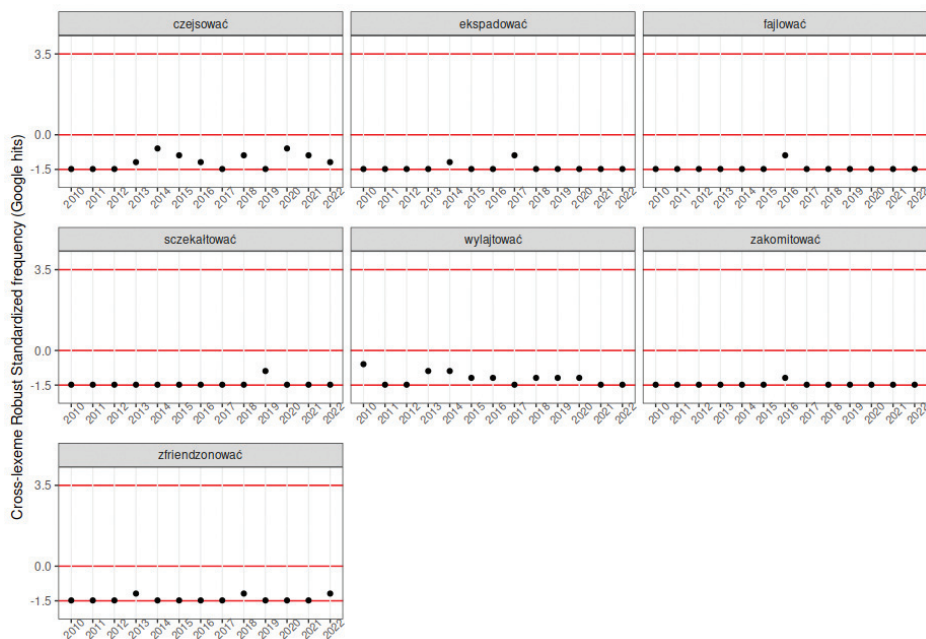
6. Discussion and implications

As was shown in the previous section, the robust z-score-based method of measuring diffusion presented in this research resulted in the creation of a five-element classification of English-origin verbs in Polish. It has to be noted here, however, that this classification is not to be considered as identifying different categories of English-origin verbs but as marking the relevant stages of the diffusion process. To make it more precise, this research views diffusion as a process extended in time that is inherently characterized by internal non-uniformity coupled with extensive internal dynamics. Accordingly, the presence of an English-origin verb in RL is viewed synchronically as exhibiting the characteristics of a given stage of diffusion process, which itself is, however, seen as a continuum (see Figure 6, for schematic representation) extending from pre-onset (of diffusion) stage to diffused stage (disregarding withdrawal stage in the synchronic perspective, as this can only be identified if the data is considered diachronically). This is made possible due to converting the observed frequencies to robust z-scores. To illustrate it better, the robust z-scores reported in this paper were calculated for the entire group of the tested verbs across

all years for which the Goggle hits were gathered. Accordingly, the median value was determined for the whole data set by ranking the observed frequencies, disregarding which English-origin verb they characterized or which year they characterized. As such, the procedure adopted in this paper computed a single set of scores based on a single measure that was not nested in the specific time of observation or the item of observation. In effect, this allows to draw insights with respect to the synchronic and diachronic aspects of diffusion the same time.
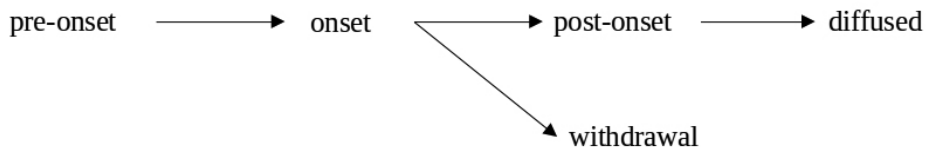

Figure 6: Proposed view of diffusion continuum

Moreover, from a diachronic perspective, the proposed view assumes a linear path of progression for a given other-language item, which is in line with the diffusion assumption (Poplack, Dion, 2012). A lone other-language item in the pre-onset stage is used by a largely limited number of speakers (perhaps even a single speaker); with time, the number of speakers using the given item increases (onset stage → post onset stage) until the item gains acceptance in a given speech community (diffused stage). However, in contrast to the previous research cited in the article at hand, the approach presented here makes an additional observation. An other-language item which begins to gain on diffusion (i.e. is at the onset stage) can develop in either one of the two directions: (i) it can increase in its diffusion even more (post-onset stage) or (ii) it can decrease in its diffusion (withdrawal). A crucial aspect of this development path is that the two potential outcomes can only occur if the initial increase in diffusion was observed in the first place, i.e. the English-origin verb has reached the onset stage. Consequently, under the view presented in this paper, other-language items in Poplack, Sankoff, Miller's (1988) or Poplack, Dion (2012) would be termed nonce type, idiosyncratic type, ephemeral type, or inveterate type have to be treated as being at a stage of their presence in RL at which it is too early to allow for any diachronic analysis.

References:

Chłopek, Z. 2011. *Nabywanie języków trzecich i kolejnych oraz wielojęzyczność. Aspekty psycholingwistyczne (i inne)*.Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.

Haspelmath, M. 2009. Lexical borrowing: Concept and issues. [In:] M. Haspelmath, U. Tadmor (eds.) *Loanwords in the World's Languages. A comparative Handbook*. Mouton de Gruyter; 35–54.

Kuźniak, M. 2009. *Foreign Words and Phrases in English. Metaphoric Astrophysical Concepts in Lexicological Study.* Wrocław: Wydawnictwo Uniwersytetu Wrocławskiego.

Kuźniak, M., Wohlfeld-Mańczak, E. 2014. Angielskie wyrazy okolicznościowe w polszczyźnie. *LingVaria IX (2014) | 1 (17)*, 69–79. doi:10.12797/LV.09.2014.17.05

Poplack, S., Sankoff, D., Miller, C. 1988. The social correlates of linguistic process of lexical   borrowing and assimilation. *Linguistics (26),* 47–104.

Poplack, S., Dion, N. 2012. Myths and facts about loanword development. *Language Variation and Change, 24*, 279–315. doi:10.1017/S095439451200018X

Poplack. S. 2012. What does the Nonce Borrowing Hypothesis hypothesize?. *Bilingualism: Language and Cognition 15 (3)*, 644–648. doi:10.1017/S1366728911000496

Poplack, S. 2018. *Borrowing: Loanwords in the Speech Community and in the Grammar.* Oxford: Oxford University Press.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

Treffers-Daller, J. 1994. *Mixing two languages: French-Dutch contact in a comparative perspective*. Mouton de Gruyter.

van Hout, R., Muysken, P. 1994. Modeling lexical borrowability. *Language Variation and Change 6*, 39–62.

Słowa kluczowe

zapożyczenie językowe, metody ilościowe, rozprzestrzenianie się, język polski, język angielski

Abstract

Feasibility of using Web-based occurrences in tracing lone items' diffusion

This article focuses on quantitative analysis of the diffusion of English-origin verbs in Polish. Based on methodology proposed by Poplack, Snakoff and Miller (1988) in their seminal work, the research presented in this paper uses frequency of occurrence obtained from Google search results to investigate the diffusion of verb of English origin in Polish between 2010 and 2022. It is shown that with proper treatment the data obtained from the Internet allow for additional insights into the process of loan items diffusion and provide basis for more fine graded classification, as well as decomposition of the diffusion process itself. The results of the analysis allow to postulate an alternative classification of loan items and capture the internal dynamics of their diffusion highlighting the significance of the initial stages of this process.

Keywords

loanwords, quantitative approach, diffusion, Polish, English