German Loanwords in Polish and Remarks on the Piotrowski-Altmann Law

Kamil Stachowski
Jagiellonian University, Cracow, Poland

Abstract. As signalled in the title, the paper has a dual purpose: it discusses the German lexical influence on Polish and also the Piotrowski-Altmann law. The first focus requires a qualitative historical linguistic approach, whereas the second is more quantitative in nature. The paper shows that when these two approaches are combined new insights can be gained.

***Keywords**: Polish, Loanwords, Piotrowski-Altmann law*

1. Introduction

Germans and Poles have been neighbours since before either became what could today be considered a nation. If one combines this fact with the consequences of the cultural and political geography of Europe, and its history, it becomes evident that the lexical influence of German on Polish must be rather substantial, and ancient. The present paper analyzes the sources (section 2) and some of the phonetic adaptations (section 3), using a primarily quantitative approach, before producing not only factual but also methodological conclusions (section 4).

Frequently mentioned is the Piotrowski-Altmann law. This is an equation which describes the progression of a change in language. The most common variant is given in eq. 1, but here a slightly modified version will be used (eq. 2; see Stachowski 2013: 110f for an explanation). The two are equivalent, with $A = ln(a)/b$.

$$P_x = \frac{c}{1 + ae^{-bx}} \tag{1}$$

$$P_x = \frac{c}{1 + e^{-b(x-A)}} \tag{2}$$

where a, $A > 0$, and b, $c \neq 0$.

The advantage of eq. (2) is that in it all the coefficients are linguistically meaningful: A denotes the moment in time when the progression of change stops accelerating and begins to decelerate (the point of inflection), b the overall speed of the change (the slope), and c its intensity (the height).

P_x itself has at least two meanings. One is the absolute count of a feature ($c > 1$), e.g. the absolute number of loanwords, as a cumulative sum. This application is used in section 2 and figs. 1 and 2. The other is the proportion of a feature ($c = 1$), e.g. of a phonetic sequence being rendered in a specific way. This

is the (attempted) application in section 3 and figs. 3–11. See also 3.2 and 4.

2. Data

There can be no doubt that German has exerted a very significant influence over Polish. Yet, to the best of my knowledge, only two comprehensive studies have been written on the subject, both quite recently: Czarnecki (2014), and WDLP. The findings of these studies are inconsistent in several regards. Let us first discuss these studies individually (2.1–2.5), take a brief detour (2.6), and only then make a more direct comparison (2.7).

The Piotrowski-Altmann curve has been fitted to both Czarnecki and WDLP's datasets and also to some of their subsets. The results are shown in Figure 1, and the coefficients are given in Table 1.

Table 1
Coefficients for the fitting of the Piotrowski-Altmann curve to the data from Czarnecki (2014) and WDLP.

Dataset	**A**	**b**	**c**	**R^2**
(a) Czarnecki	1521.28	0.01670	2594.05	0.9900
(b) WDLP	1740.70	0.06671	4048.03	0.9589
(c) ≈ Czarnecki ∩ WDLP	1547.13	0.01909	1255.97	0.9879
(d) ≈ Czarnecki ∪ WDLP	1620.48	0.00840	4404.10	0.9581
(e) WDLP main entries	2334,72	0.00425	11831.79	0.9613
(f) WDLP multiple	A_1= 1545.24 A_2= 1888.98	b_1 = 0.01876 b_2 = 0.02209	c_1= 1933.94 c_2 = 1963.68	0.9963

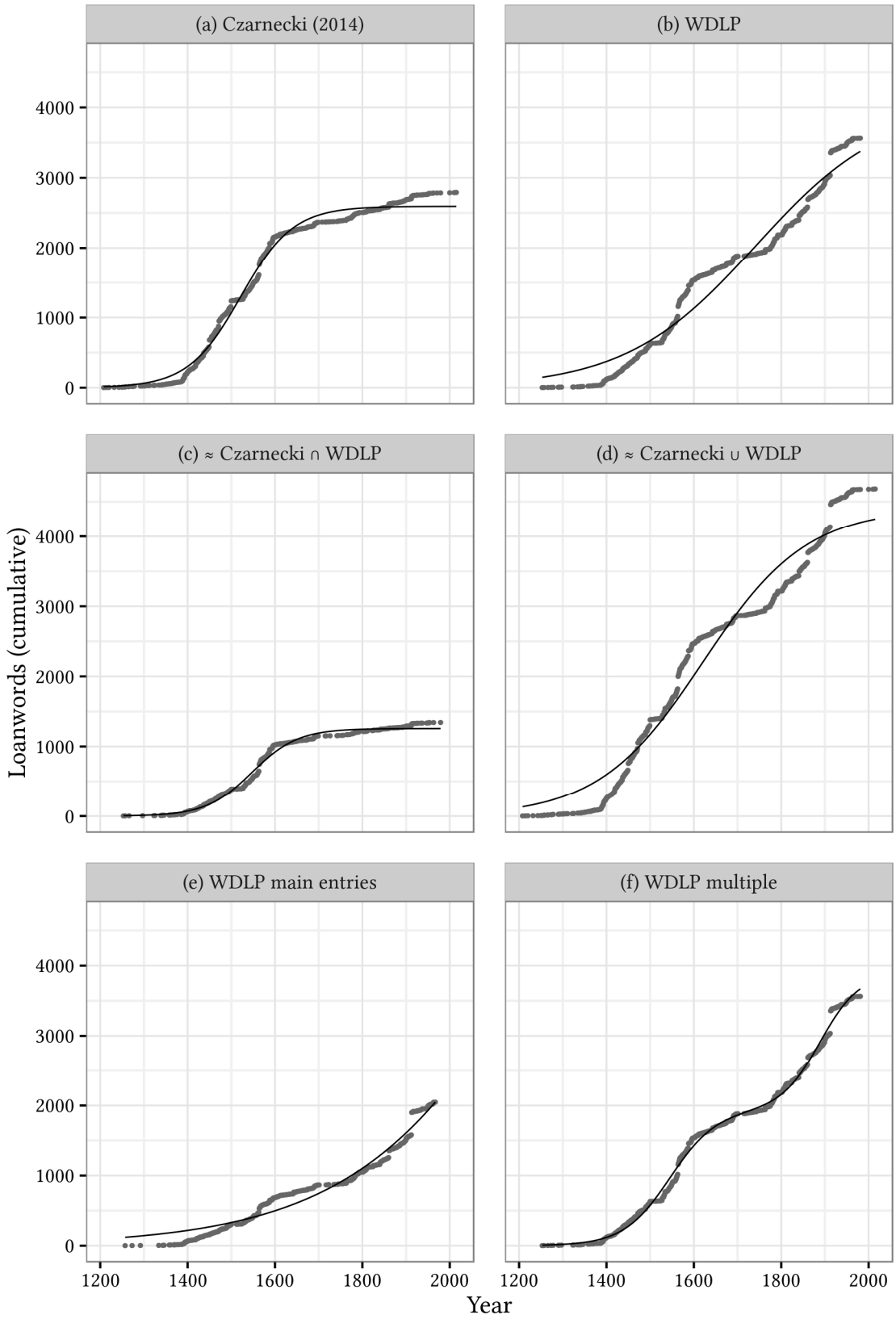


Figure 1. Fitting of the Piotrowski-Altmann curve to the data from Czarnecki (2014) and WDLP. See Tab. 1 and section 2

2.1 (a) Czarnecki (2014)

Let us begin with Czarnecki (2014). His list contains 5800 forms in total: 4936 are actual German borrowings in Polish, and the rest are either words which entered Polish through Czech mediation, Gothic loanwords, or unclear cases which may as well be native formations.

In fact, the list is a little longer because more than thirty words occur twice or even thrice in it, often with different datings and etyma, and in various places in the supposedly alphabetical order. For example, the corpus contains *szlachta*, which is dated 1390, 1378, and 1423, and derived from OHG *slahta*, MHG *slechter herr*, and MHG *slëhte* + OCz. *šlechta*, respectively (p. 267). This inconsistency is just one of the many in Czarnecki's book. Coupled with multiple unclear and unexplained cases (of which there are several examples discussed below), they force me to judge the work as unreliable. This matter will be discussed further in this section but ignored later on.

If we consider direct borrowings, the time frame is 1205–2014. The oldest words are *grenica* (1205, < MHG *grenize* or a native formation), *Roprachtovo* (1208, < MHG *Ruoprëcht*), and *barta* (1213 or 1472, < MHG *barte* rather than < OHG *barta*), whereas the newest are *Aspirin* (2014, < G *Aspirin*), *Autoland* ~ *AutoLand* ~ *Auto Land* (2014, < G **Autoland*), and *Urinal* (2015, < G *Urinal*). In this tiny sample, only *barta* is also listed in WDLP, where it is dated 1472. The form *Roprachtovo* is surprising, not least because of the use of ‹v›, a letter that is essentially completely absent from Polish orthography. In the collective list (p. 252), Czarnecki does not cite any source for *Roprachtovo* (he only cites Taszycki 1974–76 for *Ruprecht* but I could not find *Roprachtovo* in it); on p. 93, however, he cites *Jungandreas* (1928: 173) as the source for *Roprahtovo*, using ‹h› instead of ‹ch›; unfortunately, I was not able to confirm this attestation. The three newest borrowings are even more debatable. Firstly, *Aspirin* is not a common noun, as Czarnecki classifies it, but a proper name trademarked by Bayer; NKJP does not suggest any other use (I am also unaware of it) and, incidentally, dates the first usage to 1997. *Autoland* is likewise a proper name, not used in any other function, dated 1999 by NKJP, and in my opinion not necessarily a German borrowing at all (though it is, contrary to Czarnecki's asterisk, attested in German, as a proper name). Lastly, *Urinal* is also a proper name, of a medicine manufactured by a Czech company, and attested in NKJP since 2005.

Czarnecki (2014) lists German words in Polish, regardless of whether they were borrowed directly or through Czech mediation (though these are marked as such), and of whether they are ultimately of German origin or not. He also does not discern whether the word currently is or ever was included in the literary variety of Polish. As a result, many words are listed multiple times in various renderings. MHG *widerkouf*, for example, is featured no less than thirty times, as *wederkaf*, *wedyrkof*, *widerkuf*, *wyderkaw*, etc. This is not a problem, but it is a design choice that needs to be borne in mind; more on this in 2.4.

Czarnecki provides clear datings for 2792 out of the 4936 words in his list; the rest are given alternate dates, ranges, termini ante or post quos, or are only dated with precision to one century. The clearly dated words, however, conform to the Piotrowski-Altmann law exceptionally well ($R^2 = 0.99$; see fig. 1a and tab.

1). It will be important to note, in light of the coming comparison with the dataset from WDLP, that Czarnecki's data depart from the curve only slightly around year 1800, and generally appear to be a result of essentially a single period of increased influence.

2.2 (b) WDLP

WDLP's dataset consists of 5014 forms grouped into 2446 entries (more on this in 2.4). Unlike Czarnecki, the authors of WDLP chose to only include words that were borrowed directly from German to Polish *and* are native in German – a quite unorthodox decision.

The time frame is 1253–1981. The oldest words are *oberszar* (1253, < MHG *über-schar* or NHG *Überschar* or *Oberschar*), *kram* (1257, < MHG *krâm* or NHG *Kram*), and *lantwójt* (1266, < MHG *lant-voget*); the newest are *branzel* (1975 with a question mark, < NHG *Brandsohle*), *ecie-pecie* (1979, < NHG *Hätschepetsch* ~ *Hetschepetsch*), and *fakelzug* (1981, < NHG *Fackelzug*).

In this small sample, *oberszar*, *kram*, and *ecie-pecie* are confirmed by Czarnecki (2014), while *lantwójt*, *branzel*, and *fakelzug* are missing. The latter is also missing from NKJP (a Google search for *fakelzug*, however, returns several websites in Polish); *branzel* is found in NKJP since 1965, and *ecie-pecie* since 1997.

Similarly to Czarnecki (2014), WDLP lists both the literary forms of borrowed words, and their alternate adaptations which either remained dialectal or entirely hapax legomena. In a way, it too lists many words multiple times. The above-mentioned MHG *widerkouf* (spelt with a hyphen in WDLP and presented as one of two possible etyma, along with MLG *wedderkop*) has as many as 22 different adaptations. Unlike Czarnecki, however, WDLP presents Polish words using the modern orthography. This is important because, although Polish spelling is fairly historical, it no longer marks vowel length which has been present in Polish for a better part of the history of the language. Its only vestige is today the letter ‹ó› (pronounced [u] = ‹u›) while ‹á› and ‹é› are completely disused. One result of this, for example, is that the difference between *hałda* (dated 19–20th c. in Czarnecki and 1573 in WDLP) and *hálda* (dated 1573 in Czarnecki) is lost entirely. Luckily, such cases are relatively rare: there are 172 examples, out of which only eleven have clear datings.

Out of 5014 words in the WDLP dataset, 3563 are dated with precision to one year. They conform very well to the Piotrowski-Altmann law (R^2 = 0.9589; see tab. 1), but it is clear that from fig. 1b that, unlike Czarnecki's data, they represent in fact not one but two periods of increased influence. This fact has not escaped the attention of the authors of WDLP; more on this in 2.5. Interestingly, the *A* coefficient which indicates the point in time when the influx of new words stops to accelerate and then begins to slowly decelerate, falls here on the year 1741 – almost precisely on the midpoint between the two sigmoids, when the influx had actually nearly stalled in reality, only to regain momentum about half a century later.

2.3 (c)–(d) Czarnecki (2014) and WDLP

There are at least two ways in which Czarnecki's and WDLP's sets can be sensibly unified: intersection and union.

For the intersection, I chose those words which are attested both in Czarnecki (2014) and in WDLP, and have the same clear dating in both. There are only 1343 such words, i.e. less than a half of Czarnecki's clearly dated words, and not much more than a third of WDLP's. These words fall into the period 1253–1979, which is a range that is very close to either of the sources' own. They follow the Piotrowski-Altmann curve nearly as closely as Czarnecki's dataset alone, resulting in R^2 of 0.9879 (see fig. 1c and tab. 1). The division into two periods of influence, which is only lightly marked in Czarnecki's data but very clear in WDLP's, is all but invisible here.

For the union, I took the intersection as described above, plus all those words that are only listed in one of the sources, and have clear datings. There are 4683 words in total, 1343 from the intersection, 1291 from Czarnecki (2014), and 2049 from WDLP. Their dates range from the year 1208 to 2015 (the 1205 *grenica* mentioned in 2.1 above was not included because its etymology is uncertain). The more numerous contribution from WDLP appears to have dominated this set, resulting in a curve similar to that of WDLP alone (see 2.2), composed of two, clearly separate periods of influence, and R^2 of 0.9581 (see fig. 1d and tab. 1). See also 2.6 for the results of fitting the Piotrowski-Altmann curve to erroneously unified datasets.

A is the most readily readable of the coefficients in Piotrowski-Altmann law (see 1 above). I would like to point to the discrepancy between the four datasets so far discussed. According to Czarnecki's data, the turning point of German influence on Polish, the point when the influx of new words began to slow down, was in 1521. According to WDLP's data, if one ignores the double-sigmoid shape of the curve, this point occurred considerably later, in 1741. If one considers only those words on which both sources agree, the first view prevails (A = 1547), but if one takes them both at face value and examines their union, the result falls in between, in year 1620.

2.4 (e) WDLP main entries

It was mentioned in 2.2 above that WDLP groups various phonetic variants of a single etymon into main entries. This is a perfectly reasonable practice from the editorial point of view, but let us now consider what implications it has for the linguistic interpretation of a fitting to the Piotrowski-Altmann law.

Out of 2446 main entries, 2047 are dated to within one-year. The dates given by WDLP are those of the oldest variant, not necessarily the one chosen as the main entry. Let us, however, consider the reality of the borrowing. Polish literary language forms in the 16th–17th century; standard German in the 17th–18th; both began to dominate only in the 18th century; in Poland, the marginalization of dialects was nearly complete by the outbreak of World War II, whereas in Germany, the process is ongoing even today. What this means is that for a better part of the history, and therefore about two thirds of the variants listed in

WDLP, the borrowing occurred not so much from German to Polish as from one of German dialects to one of Polish dialects. Later, when those dialects produced between them the so-called literary variety, this new dialect may have or may not have inherited one specific rendering. The others remained generally restricted to dialects. Singular cases may have entered the literary language whilst it was not yet quite as rigid as it is today, and co-existed with other renderings for some time, but ultimately they too were ousted. Therefore, for those pre-literary words it is more correct to consider not the oldest variant, but only the one which eventually won out. For newer borrowings, the situation is in fact very similar.

Let us apply the Piotrowski-Altmann equation to the earliest attestations of those variants which the authors of WDLP chose for main entries. The fit is very good, $R^2 = 0.9613$ (see fig. 1e and tab. 1), but interestingly, it is actually only the initial part of the curve that covers the entire dataset. According to this prognosis, German influence on Polish is only gaining speed at the moment; the influx of new words will keep intensifying until the year 2335, and it will not slow down noticeably until around the 34th century, by which time Polish will have borrowed more than 11,000 words in total.

See 2.7 for a little commentary on the methodological question posed by this result.

2.5 (f) WDLP multiple

Let us return to considering all the attested variants, not just the ones chosen for the main entries. It was mentioned in 2.2 above that the dates given by WDLP seem actually to form not on one but two sigmoid curves, reflecting two separate periods of increased influence. The authors of WDLP have also noticed this characteristic; see especially Hentschel (2001, 2009). Indeed, having read those papers, I expected that a single sigmoid would not fit unbinned data very well. I was wrong, as is attested by $R^2 = 0.9589$. But it is also true that a sum of two sigmoids fits them exceptionally well; with $R^2 = 0.9963$ (see fig. 1 and tab. 1). A similar result was obtained with Turkic glosses in Polish (Stachowski K. 2013: 113f) and, I expect, can be obtained with the frequency with which the word *terrorismo* has appeared in end-of-year speeches of the presidents of Italy (Köhler/Tuzzi 2015: 117). All of these datasets pose a methodological question which we shall return to in 2.7.

2.6 Excursus: Garbage in, gospel out

While preparing the union of Czarnecki's and WDLP's datasets, I made a mistake and counted their intersection twice. The resulting dataset contained 6355 elements and proved to follow the Piotrowski-Altmann curve quite closely; in fact marginally closer than the result of the correctly performed union: $A = 1583.02$, $b = 0.01075$, $c = 5702.31$, and $R^2 = 0.9663$ (as compared to 0.9581, see Tab. 1).

Intrigued, I experimented with a few other datasets unified in a somewhat irrational manner, and found that the Piotrowski-Altmann law described them all with an astounding accuracy, R^2 not dropping below 0.96. An example is presented in fig. 2a and tab. 2; it is the union of three datasets: Russian borrowings

in Aleut (based on Bergsland 1994), Ottoman borrowings in Hungarian (based on Kakuk 1973; see also Stachowski 2013), and WDLP. For comparison, in fig. 2b, is a fitting of the curve to an entirely random sample of 200 numbers from the interval [1, 100]. The fit is in fact a little better, $R^2 = 0.9888$; see tab. 2. I was able to further improve it to about 0.993 (based on one hundred trials) by binning the data into 10-year intervals, and to about 0.9964 by binning into 20-year intervals.

Conclusions from this little experiment are in 2.7.

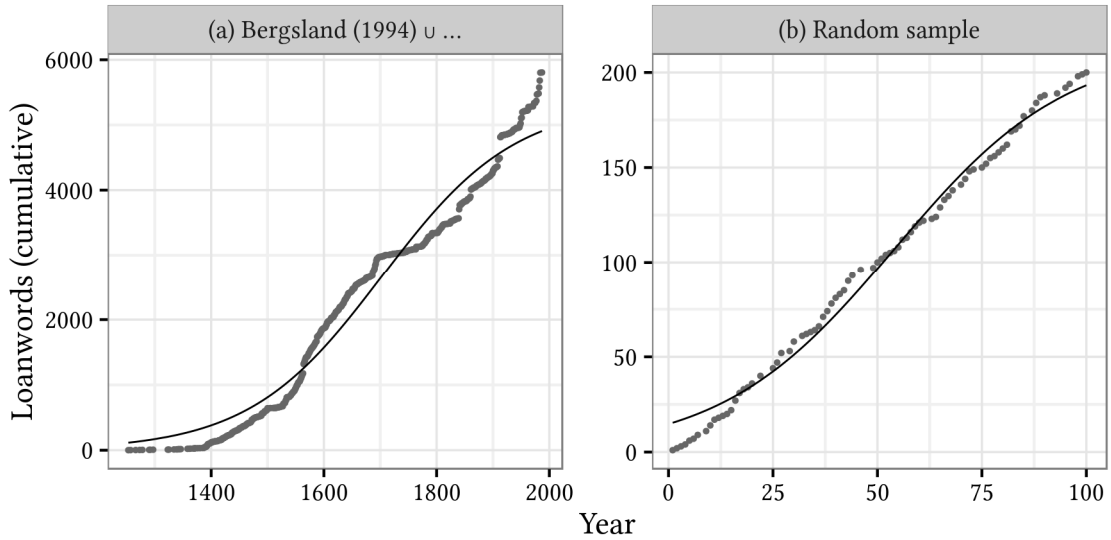


Figure 2. Fitting of the Piotrowski-Altmann curve to (a) the union of datasets from Bergsland (1994), Kakuk (1973), and WDLP, and (b) a random sample. See tab. 2 and subsection 2.6.

Table 2
Coefficients for the fitting of the Piotrowski-Altmann curve to (a) the union of datasets from Bergsland (1994), Kakuk (1973), and WDLP, and (b) a random sample. See fig. 2 and subsection 2.6.

Dataset	***A***	***b***	***c***	R^2
(a) Bergsland ∪ …	1703.57	0.00846	5351.52	0.9698
(b) Random sample	54.14	0.04818	214.48	0.9888

2.7 Observations

The first observation is about the reliability of sources (see 2.1–2.3). Czarnecki (2014), unfortunately, must be thoroughly reviewed before it can be used. Figures 1a–d serve to show the discrepancy between Czarnecki and WDLP. An interesting point is the lack of the second, (19th-century) sigmoid in the intersection of the two datasets (fig. 1c). It raises the question of why this intersection is consistent with Czarnecki's data but inconsistent with WDLP's. The two comple-

ments would need to be very carefully examined in order to formulate an answer, but, in light of Czarnecki's shortcomings, I think this may not be necessary. Nonetheless, I should like to repeat here that I find WDLP's two decisions: to present all words in the modern orthography with vowel length marking omitted, and to exclude words which are not native to German, to be quite unfortunate.

Other than that, it may be said that the two datasets contain a large number of examples (2792 in Czarnecki (2014), 3563 in WDLP) and they cover a very similar period from the 13th to the 20th/21st century, but that the pictures that they paint of the chronology of German influence on Polish are rather different. They agree about the first wave of borrowings, the one that peaked in mid-16th century, but WDLP then shows another wave culminating in the mid-19th century, which Czarnecki (2014) almost entirely omits. Indeed, to see the error in the latter view, one needs to merely remember that since late 18th century all of western and a part of central Poland was annexed by Prussia and has remained under German control till the end of World War I.

The second observation pertains more to the application and interpretation of the Piotrowski-Altmann equation than to the datasets themselves (see 2.4). It seems that one should be able to freely choose whether to include all the various phonetic renderings of an etymon, ephemeral as they may be, or to limit oneself to just the more 'successful' variants. In fact, WDLP makes a note if a specific rendering is a hapax legomenon or only attested in a single source. But the two subsets, of words with and without such annotations, both appear to be quite representative of the entire dataset in that they clearly fall on a double sigmoid, and both yield sensible results when the Piotrowski-Altmann curve is fitted to them (words not marked (2274 examples): A = 1749.76, b = 0.00633, c = 2608.28, R^2 = 0.9495; words marked (1289 examples): A = 1721.93, b = 0.00731, c = 1429.97, R^2 = 0.9657). This is not the case when only the main entries are taken into account. Visually, they may appear to fall on a curve fairly similar to the one drawn by the entire dataset, but the results of the fitting are quite different (fig. 1e, tab. 1). Only time, specifically the next millennium, can tell whether they will prove more accurate.

The third observation returns to the question of fitting multiple curves to a single dataset (see 2.5). This possibility was also mentioned in Stachowski (2013: 113f) with reference to Turkic loanwords in Polish, but the case discussed here is considerably clearer. The influx of German words into Polish can be approximated by a single sigmoid to a high level of accuracy, but it is clear from figures 1b and f that the data actually fall on two sigmoids, and this raises the question whether it should. Perhaps both approaches are valuable in their own ways. The multi-curve can pinpoint the exact moments in time when the influx of new words peaked, adding precision to a more historically-oriented perspective, while a single curve shows that even an influence which is clearly composed of two separate waves does overall follow the epidemic curve, bringing a minor new insight to the more quantitatively-oriented outlook.

The fourth and last observation will be perhaps more helpful to those who do not deal with quantitative analyses of linguistic data on an everyday basis. It was shown in 2.6 that the Piotrowski-Altmann law can describe linguistically

nonsensical data, or indeed random data, as accurately as an honest, sound dataset such as that of WDLP. One may be tempted to say 'Well, this just fits anything, then!' and disregard the idea altogether as being too broad to be able to produce an actual insight, but this would be a mistake. Firstly, the Piotrowski-Altmann curve does not in fact fit any odd bit of linguistic data, as will be demonstrated in section 3. It would also not be much use if it only fitted some of the appropriately prepared datasets but not others. Secondly, a ruler is not broken just because it can measure a broken chair. The Piotrowski-Altmann law describes how change progresses in a language, the coefficients of the regression capture numerically the time, the rate, and the intensity of the change, but a single equation cannot be expected to validate the linguistic sense of the data it was fed.

3. Method

The area of application of the Piotrowski-Altmann law is not limited to the absolute number of loanwords. Among others, it has been used with good or very good results to describe morphological changes in German verbs (Best 1983), the shift from *ward* to *wurde* in German (Best/Kohlhase 1983, Kohlhase 1983), *e*-epithesis in German verbs (Imsiepen 1983), the shift from *vi* to *ci* in Italian (Köhler/Tuzzi 2015), or the relation between grammatical markers (including fixed word order) and the number of word classes (Vulanović 2013). What these analyses have in common is that in all of them the measured value is a proportion, as opposed to a cumulative sum of absolute counts.

When reading through WDLP, it is quite evident that certain elements, sequences of sounds or parts of compounds, occur more frequently than others, and are not always rendered in the same way. I selected a dozen such elements. Let us first take a closer look at them (see 3.1), and then make some observations about the Piotrowski-Altmann law that they inspire (see 3.2).

3.1 Adaptations

Out of the dozen features selected for examination, five are elements of compounds (*-eisen*, *-haus*, *-holz*, *-meister*, *-stein*), four are sounds or sequences of sounds (‹ei›, *ke*, *l*, *VNC*), and three are affixes (*-er*, *ge-*, *-ung*). Let us look at them in the alphabetical order.

3.1.1 *‹ei›*

The graphic sequence ‹ei› appears in the etyma of 720 words in WDLP's dataset. I discarded words with significantly unclear etymology, and those in which the ‹ei› was entirely omitted (e.g. *blumistyka* < NHG *Blumisterei*, *ząząel* < NHG *Zaumseil*), or it was inside *-meister*, *-stein*, or *-eisen* as these, owing to their frequency, appear to have received special treatment, and will be discussed separately. This left 503 words containing 508 instances of ‹ei›. The adaptations attested in them are: *aj* (150 examples), *ej* (138), *y* (80), *e* (64), *a* (36), *i* (29), *u* (8), *ę* (1), *ij* (1), and *yj* (1). Of these, 333 instances are clearly dated.

Because the number of different renderings is quite high, fig. 3 shows only one aspect: whether the final *-j* of the original diphthong has been preserved. An evolution in time is clearly visible but, so far as the proportions are concerned, it hardly resembles the sigmoid of the Piotrowski-Altmann law.

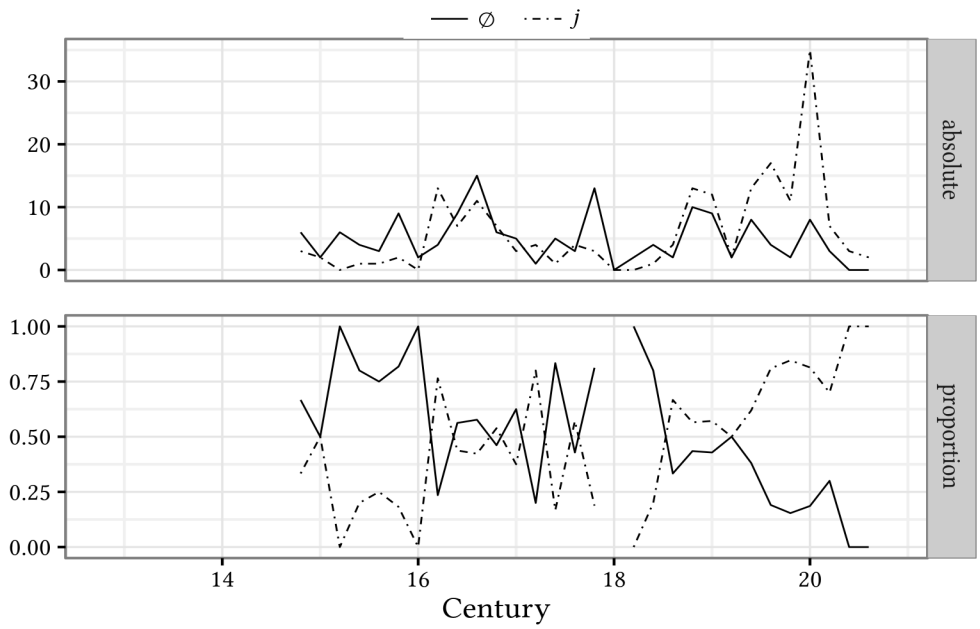


Figure 3. Rendering of *-j* in ‹ei›. Data binned into 20-year intervals. See 3.1.1.

3.1.2 *Eisen*

There are borrowings of 47 compounds with *Eisen*. I discarded three words: *arzenagiel* < NHG *Eisennagel*, *bygla* < NHG *Bügeleisen*, and *hulejza*/*hulejsa* < NHG *Hohleisen*. The adaptations in the remaining 44 are: *ajza* (18 examples), *ajz* (10), *eza* (6), *ajzen* (2), *ajzy* (2), *ejza* (1), *ejzen* (1), *ejze* (1), *es* (1), *ezyja* (1), and *yza* (1). All are clearly dated.

The set is nevertheless quite small and a graph does not reveal much more than that *ajz* is concentrated mainly around the late 18th century, while *ajza* is concentrated around the late 19th with occasional appearances in the 17th and 18th centuries. The whole dataset does not seem to follow any specific pattern. For most intervals, it is even pointless to calculate proportions because they have just one borrowing in them – or none at all.

3.1.3 *-er*

There are 845 words in WDLP whose etyma end in *-er*. I discarded those where the rendering could not be clearly established or was altogether missing (as in *demfrować* < NHG *Dämpfer*, *fercel* < NHG *Feldscher*, or *sztabstrębacz* < NHG *Stabstrompeter*), and where it was inside *meister* which appears to have been treated differently and is discussed separately. This left me with 704 examples. The adaptations in them were: *er* (377 examples), *arz* (97), *ar* (57), *ra* (37), *erz* (37), ∅ (28), *r* (20), *ir* (11), *el* (6), *irz* (6), *or* (6), *ry* (4), *y* (4), *yr* (4), *a* (2), *ro* (2), *ur* (2), *yrz* (2), *ery* (1), and *usz* (1). Of these words, 504 are clearly dated.

The primary opposition is *-er* against all the other renderings. The absolute

number of borrowings is perhaps less obvious, being distributed between the two spikes in the influx of German loanwords into Polish, but the proportion reveals the trend very clearly, as can be seen in fig. 4. It is, however, not at all similar to the curve of the Piotrowski-Altmann law.

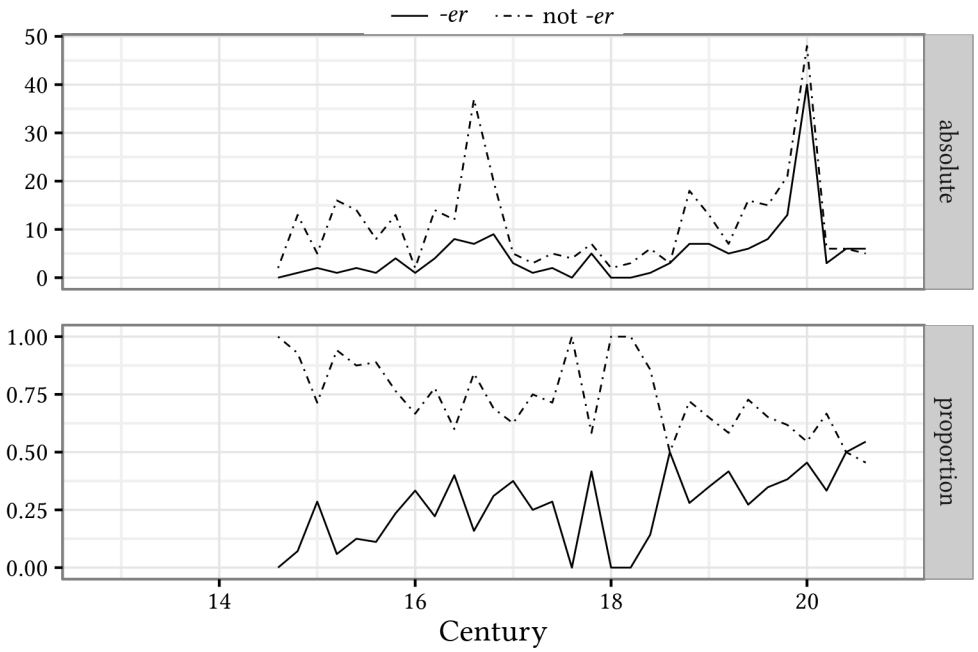


Figure 4: Rendering of *-er*. Data binned into 20-year intervals. See 3.1.3.

3.1.4 *ge-*

There are 40 words in WDLP whose etyma begin with *ge-*. I discarded those where the beginning was not the *ge-* prefix (e.g. *gierować* < NHG *gehren*, *gilować* < NHG *geilen*), and was left with 31. In 24 cases, the *ge-* was rendered as just *g-*, and in seven as *gV-* (*ge-* four times, *gie-* twice, and *ga-* once). 28 words are clearly dated.

The dataset is quite small but it has nonetheless proved sufficient for a graph: fig. 5. Assuming that the drop in *g-* at the beginning of the 16th century is accidental and meaningless, no more can be deduced from the graph than that *gV-* has effectively always been the less popular choice.

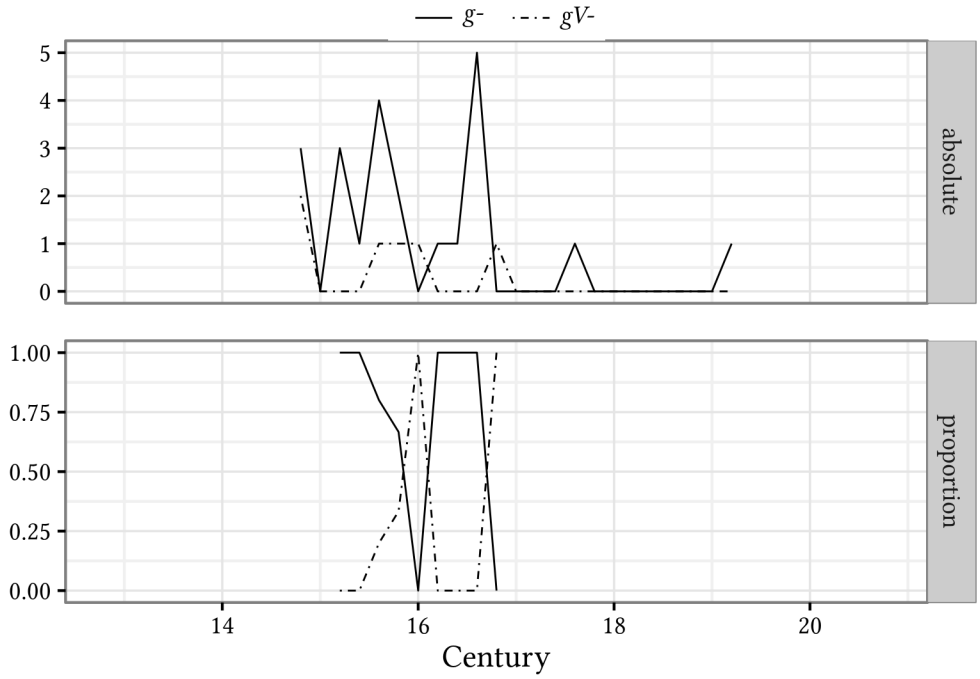


Figure 5: Rendering of *ge-*. Data binned into 20-year intervals. See 3.1.4.

3.1.5 *-ge-, -ke-*

There are 839 words with etyma containing the *g*/*k* + *ae*/*æ*/*e*/*ë* sequence. I discarded those where the sequence was entirely omitted (e.g. *bigajza* < NHG *Bügeleisen*, *ćwikać* < NHG *zwicken*, on account of *-ać* being the Polish infinitive suffix), or where it was: in auslaut, in particular as a part of the *-unge* suffix, followed by the *-er* suffix, or in the *ge-* prefix (the last three appear to have been treated differently, and are discussed separately). This left me with 322 instances in 315 words. The adaptations are: *kie*/*gie* (148 examples; labelled *K´E* in fig. 6), *k*/*g*/*h* (91; labelled *K*), *ke*/*ge* (55; labelled *KE*), *ka*/*ga* (16), *ki*/*gi* (not followed by *e*; 7 examples), *ko* (2), *go* (1), *he* (1), and *że* (1). Of these, 206 are clearly dated.

In the interest of readability, both data and adaptations are binned in fig. 6 (see the paragraph above), and only the top three groups of renderings are included, which comprise 92.7% of the clearly dated examples. In truth, the graph does not seem to clarify much. During the first wave, *K* and *KE* appear to have been almost equally popular while during the second, *KE* dominated all the other renderings. None of the adaptations even approaches the curve of the Piotrowski-Altmann law.

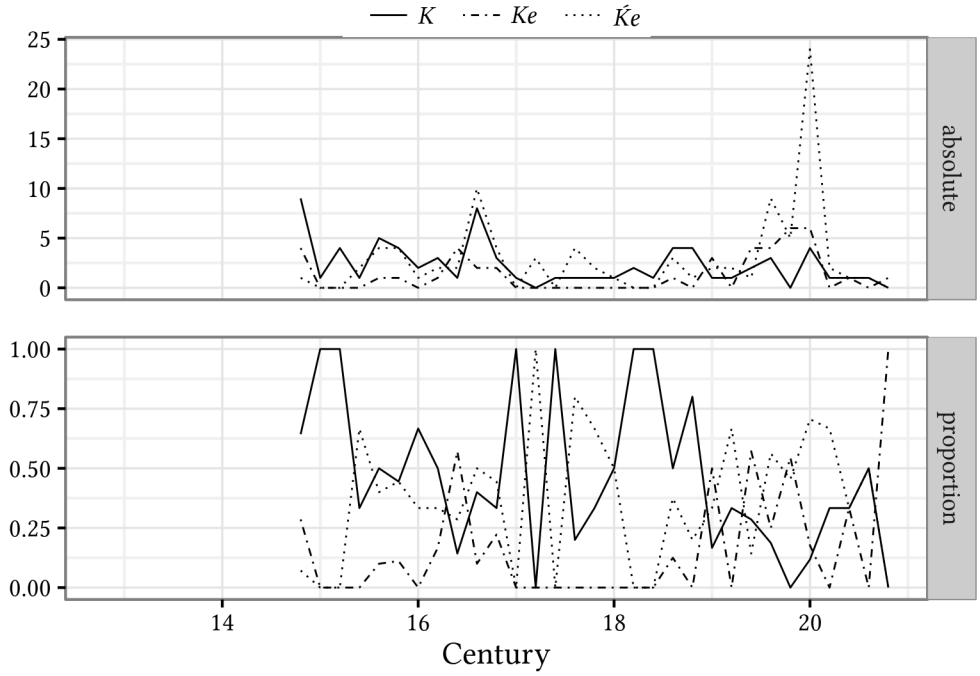


Figure 6: Rendering of *-ke-*. Data binned into 20-year intervals. See 3.1.5.

3.1.6 *Haus*

There are 88 borrowings of compounds with NHG/SilG *Haus* or MHG/MLG *hûs*. None needed to be discarded (including the nine where *Haus* was the first part of the compound). The renderings are: *hauz* (24 examples), *uz* (15), *auz* (14), *haus* (9), *us* (7), *aus* (5), *usz* (4), *hus* (2), *huz* (2), *ans* (1), *ausz* (1), *has* (1), *os* (1), *uza* (1), and *uż* (1). 57 words are clearly dated.

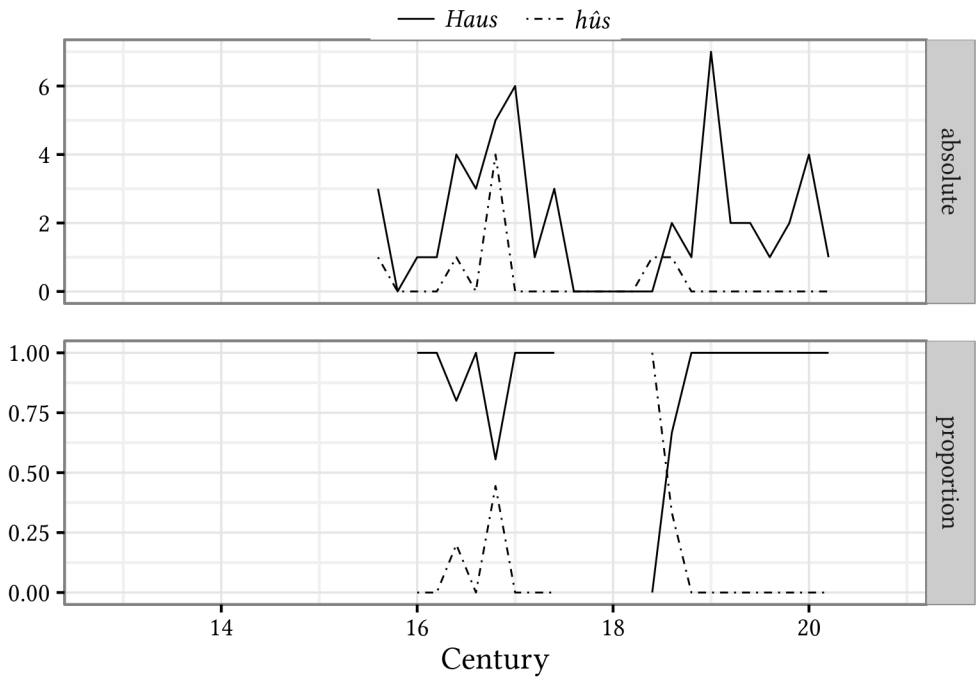
This is not a large dataset but the adaptations are quite diverse and there are too many to be legible in a single graph. It is clear that they are concentrated in two periods, the first ranging from the mid-15th to the mid-17th century, and the other from the early 18th century to the early 20th. This is consistent with the general picture painted by WDLP's data. In both periods, the different renderings are distributed approximately evenly, except for *hauz* becoming a little more prominent in late 19th and early 20th century.

Graphs of single features, such as the preservation of *h-*, rendering of the vowel, or of the final consonant, do not appear to be any more informative. Perhaps the vowel is the most interesting. WDLP reduces all the loanwords to just two etyma, MHG/MLG *hûs* and NHG/SilG *Haus*. One might expect that the first would be most prominent in the earlier wave, the second in the later one, and that the vowel in the Polish rendering would reflect that. But this is not quite the case, as can be seen from figs. 7a and b (for legibility, the latter shows only *au* and *u*, i.e. 96.5% of the clearly dated examples, and disregards the singular cases of *a*, *an* and *o*).

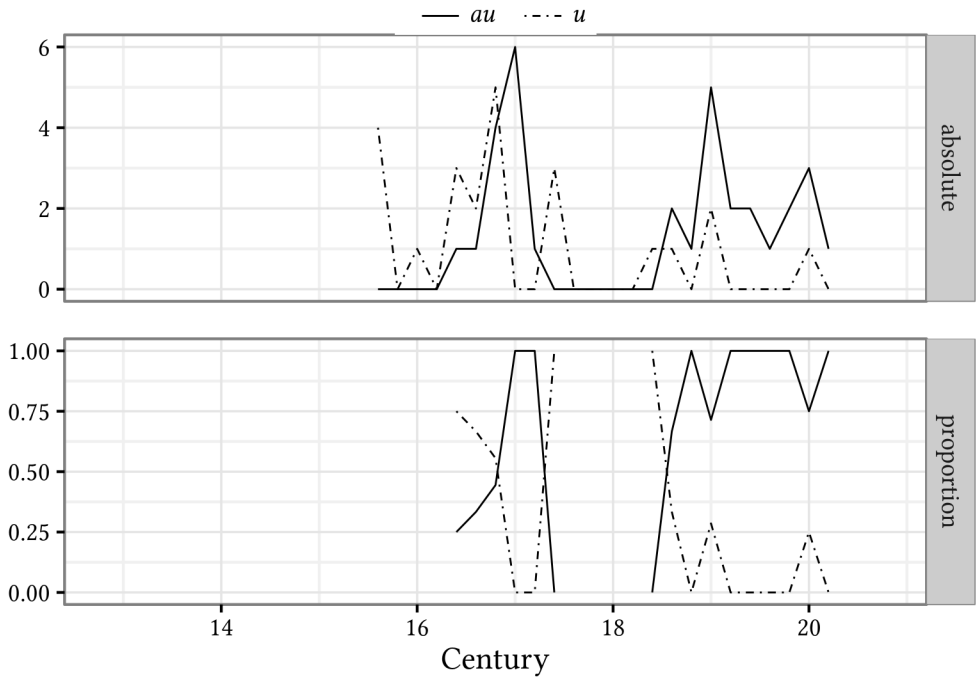
As was mentioned in 2.2, WDLP groups the various phonetic variants of a single word into entries. An etymology is only given for the entry as a whole. In

particular, an etymon with *hûs* is suggested for ten variants, nine of which are rendered with *u*. The exception is *rathaus*, dated 1767 and listed s.v. *ratusz* < MHG *rât-hûs*. It is perhaps most reasonable to view this one case as an omission on the part of WDLP, and conclude that MHG *hûs* was rendered in Polish with *u*.

With *Haus*, however, the situation is less clear. There are 78 phonetic variants derived from *Haus* (including both those with and without a clear dating); 52 are rendered with *au*, 23 with *u*, and one each with *a*, *an* and *o*. The twenty-three rendered with *u* are grouped into six entries. It does not seem very likely that they are all omissions such as *rathaus*. Perhaps they were influenced by the knowledge of previous borrowings of *Haus* that had been rendered with *u*? See also 3.1.9.



(a) The etyma according to WDLP.



(b) Rendering of the vowel

Figure 7. Adaptation of *Haus*/*hûs*. Data binned into 20-year intervals. See 3.1.6.

3.1.7 *Holz*

There are 49 borrowings of compounds with *Holz*. I discarded those where it is unclear exactly which part of the Polish word is the rendering of *Holz*, or it is absent altogether; e.g. *strycholec* < MHG *strich-holz*, NHG *Strichholz*, *Streichholz* (the *ch* may be from *stri<u>ch</u>* or *<u>h</u>olz*), or *wałek* < NHG *Walkholz*, *Walkenmangel*. The five in which *Holz* was the first part of the compound, I preserved. This left me with 38 examples, in which the following renderings are attested: *ulec* (20 examples), *holc* (11), *olc* (3), *holec* (1), *hulc* (1), *olec* (1), and *ólc* (1). Of these, only 27 are clearly dated.

Very little can be said based on this small dataset. *holc* is the most frequent during the first wave in the 15th century, while the second wave, from mid-18th to early 20th century, is dominated by *ulec*. No real patterns can be seen.

3.1.8 *l*

There are 1992 words containing 2002 instances of *l* in their etymon. I did not discard any. In 1895 examples, the rendering was *l*, and in 107 it was *ł*. 1406 instances are clearly dated.

The fairly high proportion of *ł* that can be observed in the initial phase in fig. 8 is probably accidental. Overall, the number of borrowings with *ł* remains fairly constant, and it seems that it is only due to the generally low number of loanwords in this period that *ł* happened to come to relative prominence. A similar spike in the proportion can be observed in early 18th century. If any pattern is to be seen here, it is close to constant.

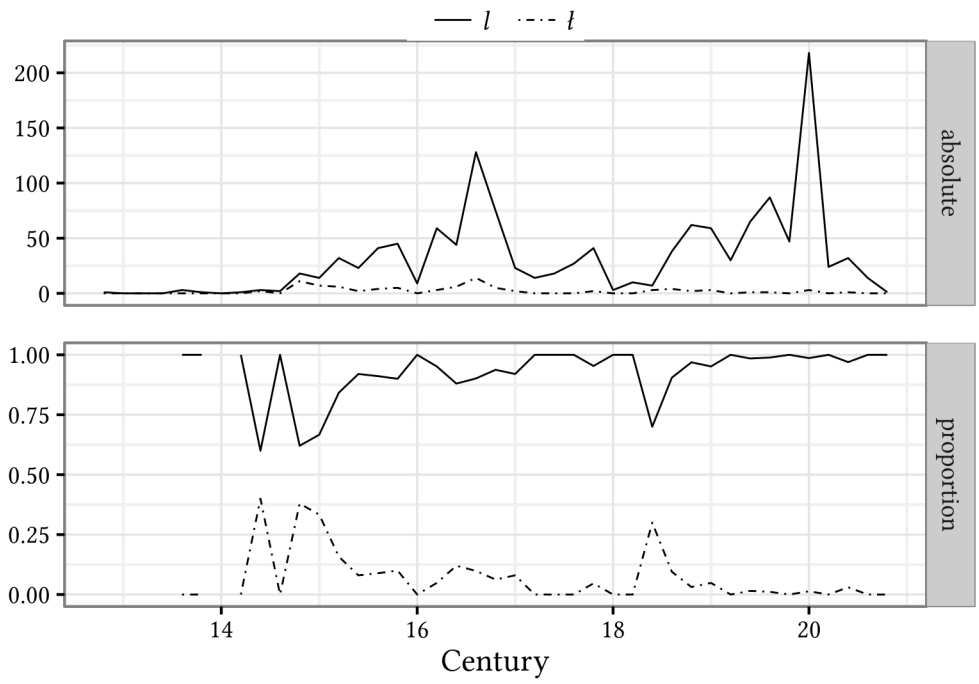


Figure 8. Rendering of *l*. Data binned into 20-year intervals. See 3.1.8.

3.1.9 *Meister*

There are 90 borrowings of compounds with *Meister*. I discarded one where this part was entirely omitted (*firtelnik* < NHG *Viertelsmeister*), and three where the spelling was unclear (*maister* and twice *meister*), but preserved seven in which *Meister* was the initial part of the compound, leaving in total 86 words. The adaptations are: *mistrz* (36 examples), *majster* (29), *mejster* (10), *magister* (6), and *mistr* (5). Of these, 57 are clearly dated.

We will start by looking at the adaptations themselves. For simplicity, I binned them into three groups: *magister*, *mAjster* (= *majster*, *mejster*), and *mistRZ* (= *mistr*, *mistrz*). Before we begin, let it be reminded that WDLP by design excludes words which are not native in German (see 2.2), and therefore all of its data and etymologies pertain to compounds containing *Meister*, but not to *Meister* as an independent word as this is a borrowing of Lat. *magister* (Kluge).

Let us then first look at *magister*. The word is missing from Boryś (2005) and Brückner (1927), but according to Bańkowski (2000), it is attested since the 15th century and originally meant 'head; manager; commander'. This is consistent with the earliest attestation in WDLP which is from year 1405, inside *och-magister* < MHG *hove-meister* or NHG *Hofmeister*. It is nonetheless quite clear that *magister* cannot be merely an adaptation of G *meister* to Polish phonetics. It must either be a calque, or borrowed from an earlier German form, one more similar to the original *magister*. The former seems to be a more likely explanation.

As for *mAjster*, etymological dictionaries quite unanimously derive it from G *Meister* (except for Brückner 1927 who is somewhat unclear), but Bańkowski (2000) and Boryś (2005) date it to the 18th century, Czarnecki (2014) to the 16th, while the oldest attestation in WDLP is from 1334, but it is inside *berkmejster* < MHG *bërcmeister*.

There is a little more uncertainty about the origin of *mistRZ*. Bańkowski (2000) derives it from OCz. *mistr* or MHG *meister* / MLG *mēster*, Boryś (2005) from OCz. *mistr*/*místr*/*mistř*, and Czarnecki (2014), together with OCz. *mistř*, from OSilG *mistr*/*mestr* (Brückner, again, is unclear). The datings, however, are more unanimous: Bańkowski to the 13th/14th century, Boryś to the 15th, Czarnecki to 1390, and WDLP to 1368, inside *barkmistrz* < MHG *bërcmeister*.

All in all, it seems that it is in fact only *mAjster* that came to Polish directly from German. Those loanwords in which *meister* was rendered either as *magister* or as *mistRZ* can be treated as borrowings of entire compounds, or maybe as borrowings of only the other part of the compound, i.e. of *och*-, *berk*- and *bark*- in the examples above, but not quite as borrowings or renderings of *meister* itself.

Nevertheless, one could hope for a pattern to emerge when the data are graphed. Fig. 9 shows the distribution, from which unfortunately no patterns seem to emerge. The overall tendencies are quite clear but the curves do not in any way resemble that of the Piotrowski-Altmann law.

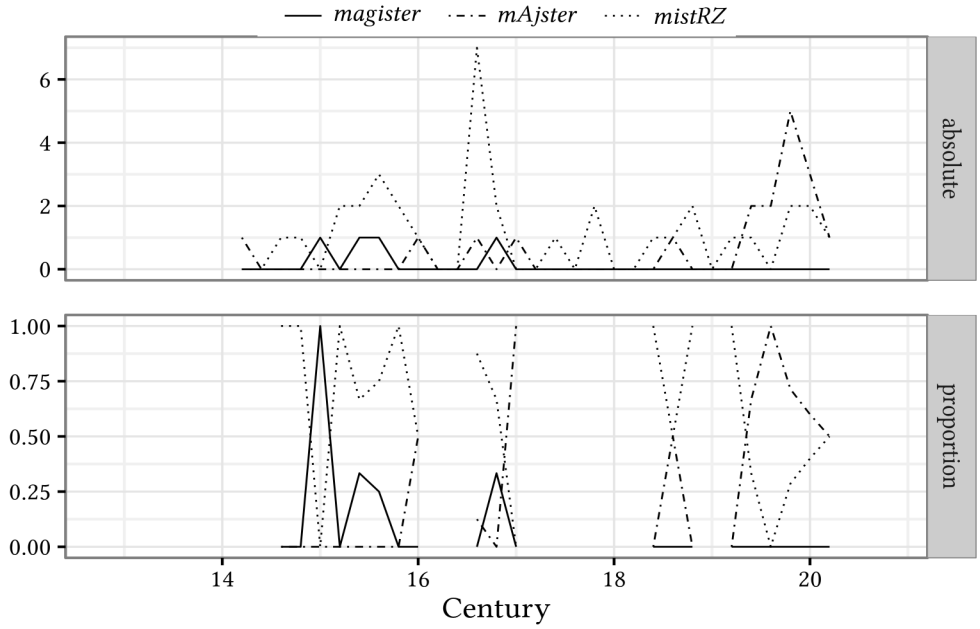


Figure 9. Rendering of *Meister*. Data binned into 20-year intervals. See 3.1.9.

3.1.10 *Stein*

There are 69 borrowings of compounds with *Stein*. I discarded those in which this part was omitted or unclear (e.g. *burztyn* < Eastern MG *bornstein*, *burnstein*, or *strab* &c. < NHG *Strebstein*) but included five cases with *Stein* as the initial part of the compound. This left me with 59 words with surprisingly diversified adaptations: *sztyn* (21 examples), *sztejn* (11), *sztajn* (10), *styn* (5), *sztyna* (3), *stejn* (2), *stin* (2), *sztan* (2), *stan* (1), *styna* (1), and *szten* (1). Of these, 32 are clearly dated.

There are too many different renderings for them to be legible in a graph, and I cannot think of a way to reasonably bin them. Graphs of single features (*s* vs. *sz* in anlaut, the middle vowel, or the gender) do not seem to be any more in-

formative. No meaningful tendencies or patterns can be observed.

3.1.11 *-ung*

There 123 words with etyma ending in *-ung* or *-unge*. I discarded those in which this part was omitted, where it was not the *-ung(e)* suffix, or the rendering was unclear; e.g. *cuhalt* < NHG *Zuhaltung*; *filun(e)k* < NHG *Füllung*; *jung* < NHG *Schiffsjunge*. This left 107 words. The adaptations are: *unek* (46 examples), *unk* (19), *ung* (13), *uga* (12), *ąg* (5), *ynek* (3), *ęg* (2), *ynk* (2), *anek* (1), *ang* (1), *onk* (1), *ónek* (1), and *unga* (1). Of these, 74 are clearly dated.

For simplicity, I binned the different renderings as follows: *UGA* = *uga*, *UNEK* = *anek*, *unek*, *ónek*, *ynek*, *UNG* = *ang*, *ąg*, *ęg*, *onk*, *ung*, *unk*, *yng*, *UNGA* = *unga*. Now, a part of the etymologies point to NHG *-unge*, a part to MHG *-ung*, and a part to both. It seems, however, that the distinction has had little to no impact on the Polish rendering: out of 19 variants with *-unge*, only one yielded *UNGA*, while the remaining eighteen are divided equally between *UNG* and *UNEK*. Out of 15 where WDLP allowed both possibilites, ten resulted in *UNEK* and five in *UNG*. The time of borrowing, it would appear, was also not decisive in any way. Fig. 10 shows the two largest groups (comprising 91.2% of the clearly dated examples) whose distribution does not follow any particular pattern.

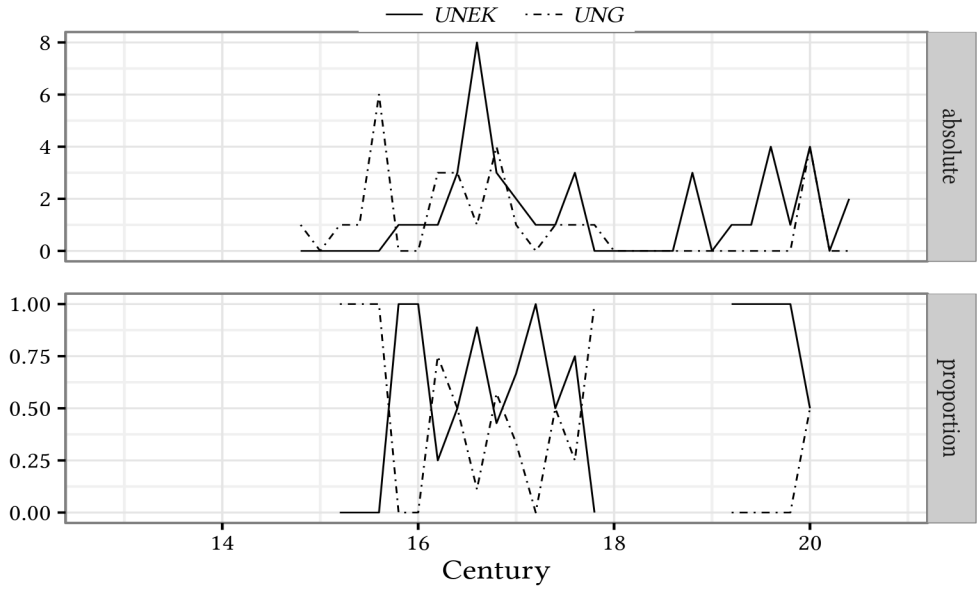


Figure 10: Rendering of *-ung*. Data binned into 20-year intervals..

3.1.12 *VNC*

There are 1268 words with the *VmC*/*VnC* sequence. I discarded those which had the sequence in an- or auslaut (based on the Polish rendering, i.e. including e.g. *antaba* < NHG *Handhabe*), where it was entirely omitted or unclear (e.g. *baleder* < NHG *Ballenleder*, or *biusthalter* < NHG *Büstenhalter*), where it was inside *Zange* used as the final element of a compound (due to their special treatment; 17 cases) or inside the *-ung(e)* suffix (likewise, discussed separately), and also 13 out of 14 different variants of *krużganek* (< MHG *kriuzganc*, NHG *Kreuzgang* or MG *krūzeganc*; the ending was always rendered as *-ganek*). This left 922 in-

stances in 898 words; in 785, the sequence was rendered more or less accurately as *Vm* or *Vn* (labelled *N* in fig. 11), in 98 it yielded a nasal vowel (labelled *Ą*), in 38 the nasality was omitted (labelled Ø), and in one case it was preserved as both a nasal vowel and a nasal consonant at the same time (*mąnsztuk* < NHG *Mundstück*).

Fig. 11 shows the top three groups comprising 99.9% of the clearly dated examples. The bulge in the proportion of ∅ in mid-15th century, and likewise the spike in early 18th century, are probably accidental owing to the generally fairly low number of examples in those periods. Apart from the general domination of *N*, hardly any tendencies or patterns are to be seen.

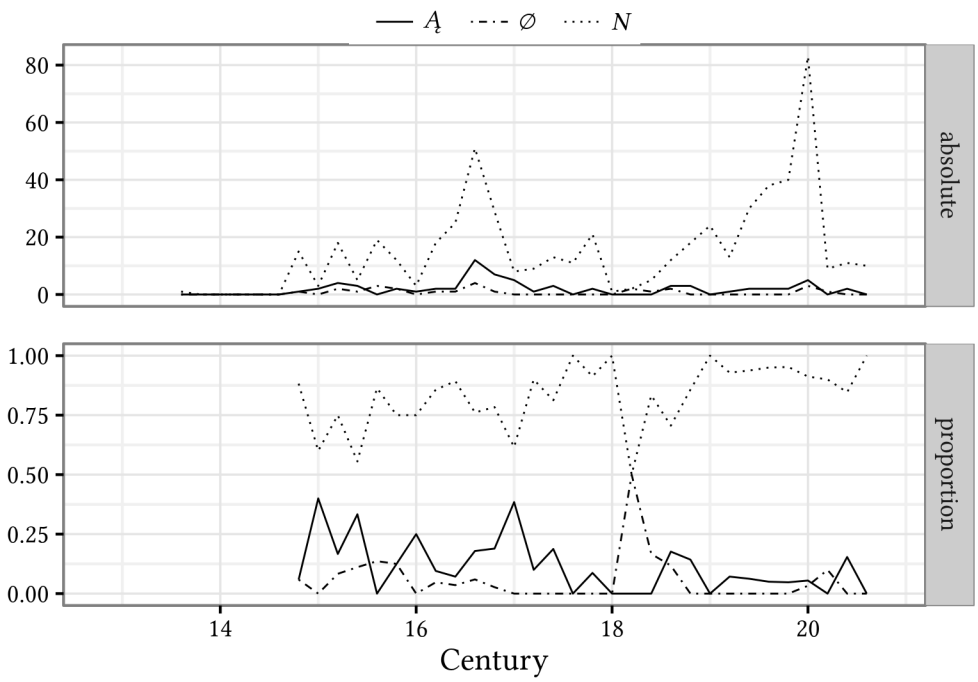


Figure 11. Rendering of *VmC*/*VnC*. Data binned into 20-year intervals.

3.2 Observations

The easiest to observe in the above analyses is the fact that the proportions of various renderings simply do not follow the Piotrowski-Altmann law. Admittedly, the situation here is not quite the same as in the several papers mentioned at the beginning of this section; what they described is the process of one form completely replacing another, while in our examples such shifts do not occur – except perhaps in 3.1.3 (fig. 4) where *-er* appears to be gradually ousting all the other renderings of G *-er*. This dataset can be approximated by the Piotrowski-Altmann equation (A = 21.51, b = 0.27888, c = 1) but only poorly, with R^2 = 0.3455. In all the other cases, the proportions do sometimes remain relatively constant, within a certain range, but mostly, they fluctuate and swap places – but without following the Piotrowski-Altmann curve. It seems that perhaps a stricter definition of the applicability of the law is required.

Naturally, if one considers the cumulative sums of absolute counts of variants with various adaptations, and limits oneself to relatively frequent renderings

(ten or more examples), the data follow the Piotrowski-Altmann curve quite closely, with R^2 ranging between 0.9076 and 0.9855. The only exceptions appear to be *holc* (see 3.1.7) and *Ke* (3.1.5) which insist on being treated as two separate sigmoids, and *mAjster* (3.1.9) which is all but missing until the mid-19th century, at which point the cumulative sum starts to grow almost exponentially. Overall, this is hardly a surprising result, since all of these datasets are more or less random subsets of a dataset which follows the Piotrowski-Altmann law with R^2 of 0.9589 (see 2.2).

The last observation is that the frequencies of various renderings appear to follow a power law distribution. Since I cannot produce a theoretical justification for this phenomenon, I limited myself to just one test using what is perhaps the most popular, the Zipf-Mandelbrot distribution. The results for adaptations with five or more renderings are promising, yielding on average R^2 of 0.9753. I expect this could be improved even further if the theoretical background allowed for a more sophisticated approach, such as in Popescu/Altmann/Köhler (2010). A report containing also examples from other languages will be published (Stachowski [forthcoming]).

4. Conclusions

German vocabulary has permeated the Polish language for more than a thousand years now, and the total number of borrowings exceeds five thousand. Despite these impressive numbers, I am aware of only two attempts at a comprehensive analysis: Czarnecki (2014), and WDLP. Both have their shortcomings, and only the latter can be considered reliable (see 2.1, 2.2, and 2.7). According to this, there have been two distinctly separate periods of increased influence. The first peaked in mid-16th century; the second, progressing slightly faster but resulting eventually in approximately the same number of borrowings, peaked in late 19th century (2.5, and 2.7).

The Piotrowski-Altmann law can be used to quantitatively describe both the whole of German influence on Polish, and the two waves in separation. It can also be used to approximate various other datasets, regardless of their linguistic significance, but one should not be discouraged by this (2.6, and 2.7). The law cannot, however, describe how the Polish rendering of various phonetic and lexical parts of German words changed over time (3.1). This demonstrates that its field of relevance may need to be defined more precisely, especially in light of the already existing dichotomy in its interpretation (1, and 3.2). On the other hand, the frequencies of these various renderings appear to be following a different law; a test fitting to the Zipf-Mandelbrot distribution yields promising results (3.2).

Abbreviations

Cz. = Czech; **G** = German; **H-** = High; **L-** = Low; **Lat.** = Latin; **M-** = Middle; **N-** = New; **O-** = Old; **Sil-** = Silesian

References

Bańkowski A. (2000). *Etymologiczny słownik języka polskiego*. Warszawa.

Bergsland K. (1994). *Aleut Dictionary. Unangam Tunudgusii. An unabridged lexicon of the Aleutian, Pribilof, and Commander Islands Aleut language*. Fairbanks.

Best K.-H. (1983). Zum morphologischen Wandel einiger deutscher Verben. In: Best/Kohlhase 1983b. 107–118.

Best K.-H., Kohlhase J. (1983a). Der Wandel von *ward* zu *wurde*. In: Best/Kohlhase 1983b. 91–102.

Best K.-H., Kohlhase J. (eds.) (1983b). *Exakte Sprachwandelforschung. Theoretische Beiträge, statistische Analysen und Arbeitsberichte*. Göttingen.

Boryś W. (2005). *Słownik etymologiczny języka polskiego*. Kraków.

Brückner A. (1927). *Słownik etymologiczny języka polskiego*. Kraków.

Czarnecki T. (2014). *Die deutschen Lehnwörter im Polnischen. Untersuchungen zur Chronologie und Geographie der Entlehnungen*. Warszawa.

Hentschel G. (2001). Das deutsche Lehnwort in der Geschichte der polnischen Sprache: Quantitäten in chronologisch qualitativer Perspektive. – Sauerland K. (ed.). *Kulturtransfer Polen – Deutschland. Wechselbeziehungen in Sprache, Kultur und Gesellschaft*. (vol. 2). Bonn. 153–169.

Hentschel G. (2009). Intensität und Extensität deutsch-polnischer Sprachkontakte von den mittelalterlichen Anfängen bis ins 20. Jahrhundert am Beispiel deutscher Lehnwörter im Polnischen. – Stolz C. (ed.). *Unsere sprachlichen Nachbarn in Europa. Die Kontaktbeziehungen zwischen Deutsch und seinen Grenznachbarn*. Bochum. 155–171.

Imsiepen U. (1983). Die *e*-Epithese bei starken Verben im Deutschen. – Best/Kohlhase 1983b. 119–141.

Jungandreas W. (1928). *Beiträge zur Besiedlung Schlesiens und zur Entwicklungsgeschichte der schlesischen Mundart im Mittelalter*. Breslau.

Kakuk S. (1973). *Recherches sur l'histoire de la langue osmanlie des XVI*^{e} *et XVII*^{e} *siècles. Les éléments osmanlis de la langue hongroise*. Budapest.

Kluge = Seebold E. (ed.). (^{25}2011). *Kluge. Etymologisches Wörterbuch der deutschen Sprache*. Berlin, Boston.

Köhler R., Tuzzi A. (2015). Linguistic Modelling of Sequential Phenomena. The role of laws. In: Mikros G..K., Mačutek J. (eds.). *Sequences in Language and Text*. Berlin, Boston. 109–123.

Kohlhase J. (1983). Die Entwicklung von *ward* zu *wurde* beim Nürnberger Chronisten Heinrich Deichsler. Als ein Nachtrag zum vorigen. In: Best/Kohlhase 1983b. 103–106.

NKJP = *Narodowy Korpus Języka Polskiego*. 2016.06.23. http://nkjp.pl.

Popescu I.-I., Altmann G., Köhler G. (2010). Zipf's law – another view. – *Quality and Quantity* 44, 713–731.

Stachowski K. (2013). The influx rate of Turkic glosses in Hungarian and Polish post-mediaeval texts. In: Köhler R., Altmann G. (eds.). *Issues in Quantitative Linguistics 3. Dedicated to Karl-Heinz Best on the occasion of his 70*th *birthday*. Lüdenscheid. 100–116.

Stachowski K. [in preparation]. A report on the distribution of phonetic rendering of loanwords.

Taszycki W. (ed.). 1974–1976. *Słownik staropolskich nazw osobowych*. (Vol. 4). Ossolineum.

Vulanović R. (2013). A multidimensional generalization of the Piotrowski-Altmann law. In: Obradović I., Kelih E., Köhler R. (eds.). *Methods and Applications of Quantitative Linguistics. Selected Papers of the 8th International Conference on Quantitative Linguistics (QUALICO). Belgrade, Serbia, April 26–29, 2012*. Belgrade. 184–193.

WDLP = de Vincenz A., Hentschel G. (2010). *Wörterbuch der deutschen Lehnwörter in der polnischen Schrift- und Standardsprache*. http://diglib.bis.uni-oldenburg.de/bis-verlag/wdlp.