Issues in Quantitative Linguistics 3

edited by

Reinhard Köhler
Gabriel Altmann

Dedicated to Karl-Heinz Best on the occasion of his 70th birthday

2013
RAM-Verlag

The influx rate of Turkic glosses in Hungarian and Polish post-mediaeval texts

Kamil Stachowski, Jagiellonian University

Abstract. The paper analyzes Turkic glosses in Hungarian and Polish post/mediaeval texts from the point of view of their correlation with historical events, and of their compatibility with the Piotrovskij-Altmann law. The correspondence is found to be very good in both cases. A slight modification is proposed to the equation to lend more linguistic significance to one of the coefficients.

0 Rationale

The goal of the present paper is twofold. On one hand, it continues the work pioneered by Karl-Heinz Best (Best/Kohlhase 1983, Best 2003, 2006, 2008, 2010 and others), of collecting empirical evidence for the so-called Piotrovskij-Altmann law. By providing Hungarian and Polish data, it also adds to the issue of Turkic influence in Europe, first discussed quantitatively in Best (2005) using the example of German.

On the other hand, it attempts to show how the quantitative and qualitative analyses complement rather than oppose or exclude each other. Luděk Hřebíček worked to demonstrate this to a more traditionalistic audience (1990: 371). My aim here is to illustrate how the quantitative approach can reveal a general tendency in a collection of detailed observations gathered and explained with the philological method.

I will: **1.** explain how I prepared the data for analysis, **2.** conduct a qualitative (philological; 2.1) and quantitative (Piotrovskij-Altmann law; 2.2) analysis, and **3.** summarize the conclusions.

1 Data

The data have been collected from two historical dictionaries (Kakuk 1973 and Stachowski S. 2007), and perhaps the majority of items are no longer in use. See 1.1 and 1.2 below for language-specific details.

Here, two general assumptions need to be justified.

The first is that I trust entirely the authors of the dictionaries in the semantic separation of items, that is to say, I always count one entry as one gloss. A precise analysis of polysemy is possible (see Levyc'kyj 2003 for an overview of methods) but beyond the scope of the present work. Since polysemous entries are rare in the material used here, I chose to assume that all such occurrences have been identified by the lexicographers and broken up into separate entries. The

second assumption is that I count each item only once, without attempting to establish whether there existed a gap in its appearance in texts. This limitation is enforced by the available data and by uncertainty as to for how long exactly it would have to be missing from the sources to be counted separately. The fact, however, that even so prepared material correlates well with the historical data on one hand, and the Piotrovskij-Altmann law on the other, appears to confirm that this disadvantage is in no way critical.

1.1 Hungarian

The Hungarian data contains Ottoman (literary and dialectal) glosses from the period 1500–1698, extracted from Kakuk 1973, which is presently the only catholic study of the subject. From a purely etymological point of view, the material is mixed in three ways and thus less heterogeneous than the Polish data (see 1.2 below).

Firstly, it is only rarely established whether an item has come directly from Ottoman. Serbo-Croatian mediation is explicitly indicated in 20 cases out of 1220 in total. Other possible intermediaries are also only sporadically mentioned. This is improbable, given that, especially in the early period of contact, neither Hungarians nor the Ottomans had good translators, and were forced to rely on foreigners, most frequently Serbians (see e.g. Hazai 1977). In the strict, etymological sense, an unknown but likely a considerable part of the material are not actually *Ottoman* glosses. In what seems to be the great majority of cases, however, their appearance in Hungarian sources can only be attributed to the cultural and administrative impact of the Ottoman rule. Thus, from the general point of view of influence, they have to be considered Ottoman glosses after all, regardless of what the immediate donor was.

Secondly, the exact Turkic source within the linguistic melting pot of the Ottoman Empire is almost never established. Etymologically, this is a significant deficiency. From the point of view of the Hungarian population, however, it would have been quite irrelevant whether a particular word is known in the entire Empire or limited to just one dialect which will, in the remote future, become Turkish or Azerbaijani, or whether it ultimately stems from Proto-Turkic or Arabic or Persian. Rather, it would have been just another word used by the occupying soldiers, settlers and tax collectors, and as such, an Ottoman gloss.

Thirdly, significant parts of the material are personal and place names which only ever occur in a Hungarian context as *Fremdwörter*. What is important for me here, however, is not how much and for how long Hungarian has made them its own, but the fact that they do appear in it at all. Even a hapax legomenon is a proof of interest and, indirectly, of influence. (In an extreme case, perhaps even better a proof than the longevity of a specific borrowing. Else, in the particular case of Hungarian, the European influence would have to be concluded

nearly purged by the end of the 19th c.)

The above does not mean in any way that more detailed studies of Ottoman influence on Hungarian are unnecessary. It merely attempts to justify why greater precision is not a sine qua non in this case.

The total size of the dataset was 1220 items. From these, I removed in the following order those cases where the dating was:

1. imprecise because the source was published over a period of several years (32 items),
2. uncertain because Kakuk (apparently) only used a later copy of the original document (4 items, see Kakuk 1973: 11),
3. imprecise (a century given rather than a specific year; 11 items) or
4. unrepresentative of the given period because it lay outside of the primary scope of my source (16th–17th c., see Kakuk 1973: 8) (11 items in years 1405–94 + 61 items in years 1701–1873, but see 2.1.1 below).

Points 1. and 2. could perhaps be saved for our purpose by some arbitrary method, as e.g. taking the earlier date, or the appropriate fraction for each year in the range. But this would not make a significant difference: for the items in point 1., the mean of glosses per year ($\overline{P}$) = 2.46, standard deviation (σ) = 2.54; for those in point 2., $\overline{P}$ = 1, σ = 0. Points 3. and 4. cannot be included in any meaningful way.

The pruned dataset contains 1101 glosses from 169 unique years from the period 1500–1698. There are 1–56 glosses per year with $\overline{P}$ = 6.515, σ = 8.227.

1.2 Polish

The Polish data contains Turkic (Ottoman and Tatar) glosses from the period 1388–1791, extracted from Stachowski S. (2007), which also is presently the only catholic study of the subject. Etymologically, the material is more homogeneous than the Hungarian data (see 1.1 above), albeit still mixed.

Firstly, it contains items not only from Ottoman and its dialects but also from Tatar. Separating these two groups from each other is often difficult or effectively impossible because of relatively high phonetic similarity of the two languages, especially when forced through the Polish phonological filter. Hence, the data must be viewed as reflecting the linguistic influence of the *Turkic element* on Polish, rather than that of Ottoman or, the more so, literary Ottoman.

The other two limitations of the Hungarian material, namely the unknown route and a high number of personal and place names, are only marginally valid for the Polish data.

Probably, the great majority of borrowings from Ottoman (the Ottoman Empire, see 1.1 above) are direct. In the case of words from Tatar, perhaps Ukrainian or other mediation might have been more frequent. As for personal and place names, these are very rare in S. Stachowski's dictionary.

There is one more major difference between the two datasets analyzed here. Unlike the Hungarian data, the Polish material has been extracted in a greater part from works which deal very specifically with Turkic matters, such as accounts of legations to the Ottoman Empire. This has resulted in a few large skips. I will consider here both the entire dataset, and its trimmed down, more continuous subset.

The total size of the dataset was 1204 items. From these, I removed in the following order those cases where the dating was:

1. imprecise because the source was published over a period of several years (117 items),
2. uncertain because the source was published 35 years after it had been written (11 items from one source, see Stachowski S. 2007: XXV s.v. *LeszD*),
3. imprecise (half a century given rather than a specific year; 4 items) or
4. unrepresentative of the given period because it lays outside of the primary scope of my source (14^{th}–18^{th} c., see Stachowski S. 2007: XVII) (176 items in years 1812–1899).

As opposed to the Hungarian data in 1.1 above, here including the items from point 1. would make a significant difference: out of the 114 glosses, 83 come from one source, eleven from another, and only the remaining twenty are distributed evenly among 15 different sources. The former of these two sources has been written in years 1496–1501 (see Stachowski S. 2007: XXV s.v. *KT-Z*), the latter – in years 1582–84 (see p. XXIX s.v. *RadzPZŚ*). Unfortunately, there seems to be no objective and good way to interpret these cases.

Items from point 2. would not make a significant difference. Items from points 3. and 4. cannot be included in any meaningful way.

The pruned dataset contains 896 glosses from 130 unique years from the period 1388–1791. There are 1–203 glosses per year with $\overline{p} = 6.9$, $\sigma = 20.59$ and kurtosis (Γ_2) = 65.56.

As the distribution of items per year is extremely uneven, I will also use here a trimmed down subset, from which the eight most bountiful years (sources) have been removed. (They supply, in order, 203, 78, 58, 51, 44, 43, 29 and 24 glosses. The following years all bring less than 17.) This subset contains 366 glosses from 122 unique years from the same period as the full set (1388–1791), with $\overline{p} = 3$, $\sigma = 3.29$ and $\Gamma_2 = 4.34$.

2 Analysis

I will give a qualitative (philological) and a quantitative (Piotrovskij-Altmann law) analysis. The former suggests that there exists a correlation between the influx rate of unique glosses and historical events. The latter, that this influx can be modelled by the equations given below.

2.1 Qualitative (philological)

2.1.1 Hungarian

Fig. 1 shows the influx of Turkic items in Hungarian texts in years 1500–1757. This is a slightly longer period than the focus of Kakuk 1973 (see 1.1 above), and the data from the final 58 years must be considered unrepresentative. I chose to include them here nonetheless because of one important remark that needs to be made about them.

The period of Ottoman rule in Hungary, 1541–1699, is hatched in the figure. Vertical lines represent what appear to be the most important and/or relevant historical events of the time (based mostly on Kálmán 19s89 and Molnár 2001). They are:

1521: Belgrade is conquered by the Ottomans.

1526: The Hungarian army is defeated at Mohács. The Ottomans enter and pillage Buda but retreat soon afterwards, holding however central Hungary and suzerainty over Transylvania.

1541: Buda is conquered; central and southern Hungary is annexed.

1547: A five-year armistice is concluded in Edirne.

1570: The semi-independent Principality of Transylvania is established and will last until 1711.

1683: The Ottoman army is defeated at Vienna, leading to the liberation of Hungary during the next sixteen years.

1686: Buda is reconquered by the Holy League's army.

1699: A peace treaty is signed in Karlowitz, effectively transferring Hungary to the Habsburgs.

1718: A peace treaty is signed in Passarowitz, transferring the Banat of Temeswar and much of present-day Serbia to the Habsburgs.

1739: A peace treaty is signed in Belgrade, restoring Ottoman administration in Serbia, the southern part of the Banat of Temeswar and northern Bosnia.

Rather surprisingly, the conquest of the country, the greater part of which had happened till around 1550, appears to have had next to no impact on the influx of Turkic vocabulary. A pronounced rise can be seen shortly after 1550, probably due to a more wide-spread application of Ottoman administration. Another surge occurs after 1570, perhaps following the establishment of the Principality of Transylvania which remained through most of its history under Ottoman suzerainty.

The small fluctuations around 1615 and 1650 are difficult to explain by historical events. Given the modest scale, they can probably be attributed to the random factor.

After 1660, the influx rate visibly drops. Like the two aberrations above, it, too, does not seem to coincide with any particular historical event, but unlike them, this is a long-term phenomenon. Perhaps then, it is to be concluded that

after well more than a hundred years of close contact, Hungarian has finally reached a (temporary) point of saturation.

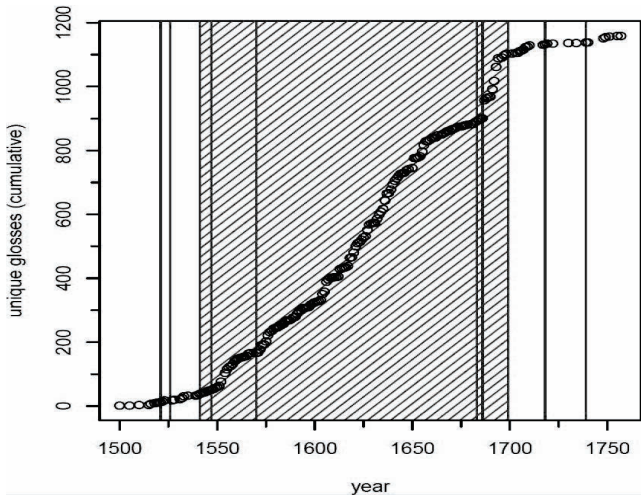


Figure 1. A cumulative sum of unique Ottoman glosses in Hungarian texts in years 1500–1757. Hatched is the period 1541–1699; see 2.1.1 for the meaning of vertical lines.

The situation changes dramatically during the War of the Holy League (1683–1698), and specifically after the reconquest of Buda in 1686. Unsurprisingly, the hope for regaining the country must have renewed the interest in Turkic matters which had lasted almost throughout the war, until the Peace of Karlowitz (1699); see also 2.1.2 below.

Subsequent events of the war between the Habsburgs and the Ottomans in present-day Croatia, Bosnia and Transylvania apparently left Hungarians indifferent, but see below. It is to be noted, however, that even after the end of the Ottoman occupation, new Ottoman items still continued to appear in Hungarian texts. This might seem counter-intuitive at first but can be easily explained. First, these items are glosses, not necessarily actual loanwords. Though weakened, the Ottoman Empire did not cease to play the role of a superpower and as such, to attract attention. Also, it is natural to expect historical treaties dealing with the period of occupation, to begin to appear after it had ended, and to contain at least a small number of previously unattested lexical items.

To sum up, the change of the influx rate can, in most cases, be explained by historical events. Interestingly, however, the coincidence does not hold equally well the other way round. 18th c. data are uncertain because this period lies out-

side of the focus of Kakuk 1973 and the numbers might be too low. Particularly surprising, however, is the early period of conquest until around 1550, especially if contrasted with the striking surge after the liberation of Buda in 1686. One conceivable explanation is that Hungarian authors would have been more eager to write about what they probably saw as a success than what obviously was a failure. Further study is necessary to provide a more certain answer.

2.1.2 Polish

Fig. 2a shows the influx of Turkic (Ottoman and Tatar) items in Polish texts in years 1388–1791. Hatched are the periods of big wars; vertical lines represent what appear to be the most important and/or relevant historical events of the time (based mostly on Davies 2005 and Markiewicz 2002). They are:

1444: The Hungarian-Polish army is defeated at Varna. The consequences are much more significant for Hungary than for Poland.

1485–1503: The First Polish-Ottoman War. The contact with the actual Turkic element, however, is limited.

1533: A peace treaty is signed in Istanbul, valid for the life of both rulers.

1569: The Union of Lublin replaces the previous personal union and creates a single state, the Polish-Lithuanian Commonwealth.

1591–93, 1594–96: The Kosiński and Nalyvaiko Uprisings, which both begin more as private quarrels but soon transform into civil wars between the local nobility and the Cossacks.

1620–21: Another conflict with the Ottoman Empire, which begins with a defeat of the Polish-Lithuanian army at Cecora in 1620 and ends with the indecisive battle of Khotyn in 1621 and the subsequent, equally ambivalent peace treaty.

1633–34: Another conflict, concluded eventually by a peace treaty in 1635 which effectively extends the previous, 1621 peace.

1648–55: Khmelnytsky Uprising, which begins more as a peasant revolt but soon transforms into a Ukrainian war of liberation of a sort; the offending party are primarily Zaporozhian Cossacks, supported by Crimean Tatars who, however, change sides twice.

1672–76: The Second Polish-Ottoman War. In 1672, the Commonwealth signs an unfavourable peace treaty in Buczacz. The war concludes with the Treaty of Żurawno in 1676 which only revises a few points of the 1672 peace.

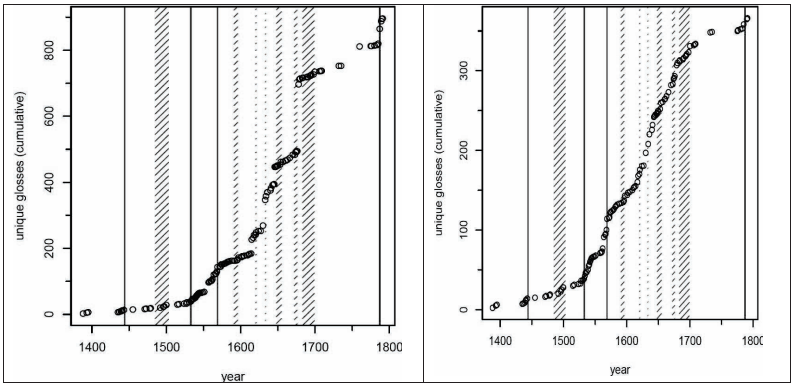
1683–99: The Third Polish-Ottoman War, which starts with the epic victory at Vienna in 1683 and ends with the Treaty of Karlowitz in 1699.

1787: The Russo-Turkish War breaks out, raising Poland's hopes of loosening the Russian protectorate.

Fig. 2b shows the same historical events as fig. 2a but superimposed on it is the trimmed down, more continuous subset of the Polish data (see 1.2 above).

The decimation of the Hungarian-Polish forces at Varna does not seem to have had much impact on the influx of Turkic glosses into Polish sources. A

slight rise is visible before the actual event. This is only mildly surprising. Overall, the aftermath of the battle was very significant, as it in fact paved the way to the Constantinople for the Ottoman Empire, but it was not critical for Poland.



2a. The entire dataset. 2b. The trimmed subset.

Figure 2. A cumulative sum of unique Turkic glosses in Polish texts in years 1388–1791. Hatched are the periods of big wars (1485–1503, 1591–96, 1620–21, 1633–34, 1648–55, 1672–76 and 1683–99); see 2.1.2 for the meaning of vertical lines.

More unexpectedly, the First Polish-Ottoman War, too, has only left a small imprint on the number of glosses. It is true that the war was relatively short and, in essence, consisted of two battles, but both were big and important ones. Perhaps it was the Lithuanian-Russian War (1512–22) and the Polish-Teutonic War (1519–21) that had drawn the attention away from the conflict in the South.

The influx rate appears to finally rise after 1533 when a long-term treaty had been signed between Poland and the Ottoman Empire, guaranteeing many years of peace. It might be suspected that this encouraged trade and other forms of pacific contacts, which caused the said rise.

Around 1550, a fluctuation can be seen, which does not seem to coincide with any significant historical event. Possibly a random variation? The clear slowdown after 1569 should probably be attributed, again, to the shift of public attention away from Turkic matters.

The first bigger difference between figures 2a and 2b turns up shortly before 1600. Inflated by a single source (Rycaut/Kłokocki 1678 with 203 items), the year 1678 has visually dominated the entire plot. The momentary increase at that time, possibly due to the two subsequent Cossack uprisings and the general tension in contacts with the Crimean Khanate, is barely visible in fig. 2a. Therefore, as far as a qualitative analysis is concerned, I believe that a somewhat

random occurrence, which the publication of one description of a country in 1678 surely is, is better disregarded if it is to eclipse a valid coincidence with historical data.

In 1620, the first of a series of consecutive wars breaks out. Shortly before it, the influx visibly regains momentum, which it is to maintain throughout the wartime, till the end of the century. It needs to be noted that in the figures, only the most important wars are marked, while in reality the entire 17th c. was defined by almost unceasing conflicts with the Ottomans, raids to and from the Crimean Khanate and Cossack uprisings (not to mention the hostilities with Sweden, Russia, Moldova and others).

Although somewhat random in nature, the skips in fig. 2a coincide surprisingly well with the periods of (big) war. In years 1645–76 and 1679–1709, however, this figure shows a slowdown whereas in fig. 2b, the rate is almost constant. A similar situation happens in year 1760 where the number of glosses rapidly increases but, as can be seen from the two figures, inflated by a single source. Again, the culprits here are the few unusually bountiful sources which distort the general picture.

Finally, in 1787 the last surge begins. In opposition to the previous two, this one has been brought about by multiple sources. Most probably, it is to be connected with the breakout of the Russo-Turkish war in that same year, which had kindled a hope in Poland that some loosening of the Russian protectorate over the country would soon follow.

To sum up, most changes in the influx rate can be explained by historical events which, in turn, are rather well reflected in the linguistic material significantly better than was the case with Hungarian in 2.1.1 above. An interesting observation is that the peace of 1533 appears to have caused an equally sharp boost, if less long-lived, as a century of almost permanent conflict.

It is not clear to me, which of the figures 2a and 2b gives a more appropriate overview of the course of influence. The former is jagged, as would be expected to result from a series of individual contacts. But on the other hand, these contacts were very numerous and followed very closely one after the other, almost blending into one, long period of war. It is perhaps debatable, what *random* actually means in our context.

2.1.3 Conclusions

In the two cases analysed here, the coincidence between historical events and linguistic data has proven to be quite accurate and mutual, that is, to be a *correlation*. This is key, as it not only provides empirical grounds for intuitive explanations resorted to in linguistics, but also it paves the way for quantitative linguistic data to support historical interpretations.

Two interesting conjectures can be made: 1. a historical success is more likely to cause a surge in the number of glosses than a failure (based on 2.1.1),

and 2. an increase in the number of glosses is equally likely to be caused by hostilities as by friendship (base on 2.1.2). Both still require further study.

2.2 Quantitative (Piotrovskij-Altmann law)

The Piotrovskij-Altmann law states that change in language can be modelled with the logistic equation (1):

$$(1) \qquad p(t) = \frac{c}{1 + ae^{-bt+Ct^2}}$$

where *p(t)* is the number of forms in question at time *t*, and *a*, *b*, *c* and *C* are coefficients.

The law has three variants which describe: 1. a complete change (the replacement of an obsolete form by the new one, where $c = 1$ and $C = 0$), 2. a partial change (where $C = 0$, see below), and 3. a reversible change (one that ended before it could dominate the language). On the history of the law, see e.g. Altmann (1983: 59–60), Lehfeldt/Altmann (2003: 142–44) and Leopold (2005: 627–28, 631–32), but also Vulanović (2007: 112f, 116f, 120).

Our point of interest here is the middle variant (there is no limit to how many glosses can possibly appear in texts, and once they appear, they become attested and cannot be taken back), which is given by

$$(2) \qquad p(t) = \frac{c}{1 + ae^{-bt}}$$

The function is sigmoid (has the shape of a stretched letter *s*) and fine-tuned by the three coefficients which control: a (> 0) – the horizontal displacement of the centre of the slope, b ($\neq 0$) – the steepness, and c ($\neq 0$) – the height. In other words, it states that a change begins slowly, then accelerates in its middle phase and later slows down again as it nears its end. It will be shown below that the intuitiveness of this observation is, to some degree, illusory.

As can be seen, the only independent variable is time (t). Additional, process-specific parameters, such as the duration or intensity of linguistic influence, are encoded in the coefficients and calculated anew for every analyzed change. Thus, the coefficients can be viewed as indices of a kind: a describing the point in time at which the process intensifies, b – its intensity, and c – its strength (the degree of influence). This is illustrated in fig. 3. Note that a and b depend on the employed time scale and are meaningless if the unit of time is not specified (year, half a century &c.).

The point of the analysis is to find how well the curve given by eq. 2 can approximate the empirical data. The goodness of fit is typically measured by the R^2 coefficient of determination, with the maximum value of 1 which defines a perfect fit.

As advised by Best/Beöthy/Altmann (1990: 117), the data are usually grouped into periods, typically of twenty-five, fifty or a hundred years. Randomness is thus partially removed from the picture and the resulting plot clearer. However, some potentially interesting details may also be lost. I chose to be as precise here as the data allow, which is one year.

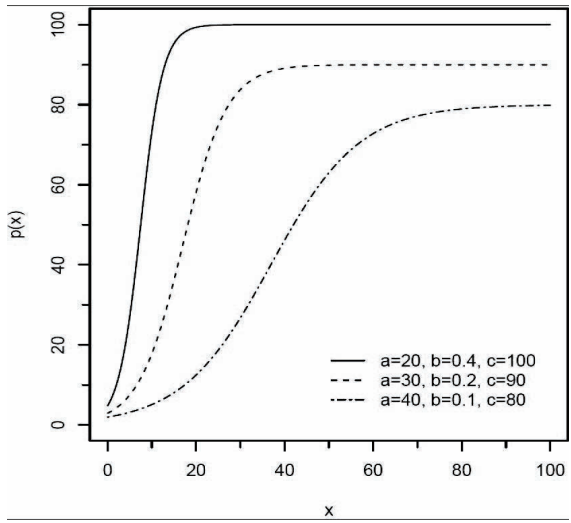


Figure 3. A comparison of function *p* (eq. 2) with different coefficients.

The numbering of years and centuries is usually transformed to begin with 1. It is, after all, purely contractual, and large numbers tend to hinder the fitting because of coefficient *a*. However, my goal here is to demonstrate how the qualitative and quantitative approaches can be used together. Leopold (2005: 631) classifies our law as an *intuitive Heuristik, die allerdings der empirischen Überprüfung standhält*. The parameter *a* does not represent any particular linguistic, historical or other value. Hence, it seems potentially beneficial to transform the equation to

$$p(t) = \frac{c}{1 + e^{-b(t-A)}} \tag{3}$$

where $A = ln(a)/b$ and, unlike *a*, remains within sane limits even with higher year numbers (e.g. with the Hungarian data below, $A = 1633.038$ which corresponds to $a = e^{47.298} = 3.476 \times 10^{20}$ in eq. 2). Also, *A* is much more meaningful linguistically, as it marks the precise centre of the sigmoid, i.e. the turning point of the influx rate.

2.2.1 Hungarian

Fig. 4 shows the Hungarian empirical data as points and the fitted curve (eq. 3, fitted with *R*). The fit is quite good, with $R^2 = 0.9924$.

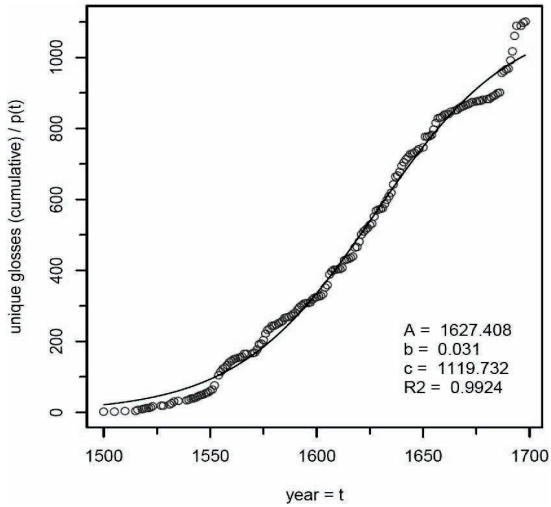


Figure 4. Fitting eq. 3 to the Hungarian data.

Perhaps in most linguists' intuition, the slow-quick-slow scheme associates with a single period of intensified contact, such as a war or occupation, and every another such period would repeat the same pattern from the start, in complete detachment from the previous one or ones. The Hungarian data show that this is not necessarily true.

Firstly, there are in fact two sigmoids visible in the empirical data, and the greater parts of both fall entirely within the one period of occupation (fig. 1; see also 2.2.2 below). Secondly, the two separate sigmoids can still be approximated quite accurately with just a single curve (fig. 4). Neither does one period necessarily cause one sigmoidal influx, nor are the subsequent sigmoids in complete detachment from each other. This seems to prove that the Piotrovskij-Altmann law is less intuitive than it might seem at face value.

These observations would be impossible if years were grouped into intervals of e.g. 25, as has often been the custom in quantitative studies. The appropriate data are given and plotted in fig. 5.

The question, then, is how much precision is optimal. The second sigmoid in fig. 4 might be an irrelevant, random fluctuation, which grouping only irons out – or a perfectly valid occurrence, which would have contributed a deeper

insight, had grouping not obscured it. But greater precision might also be misleading. Typically, sources used for historical dictionaries are prints. In each case, the time that elapsed between the moment when the original text had been written, and the year when it was printed, is different. Grouping helps remove this randomness from the picture. On the other hand, the surge in 1686 can be rather believably explained by the reconquest of Buda. Without sound empirical evidence, this discussion appears to be a battle of beliefs.

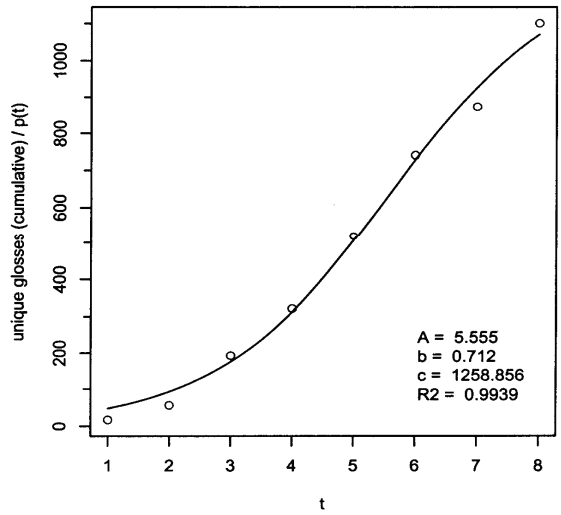


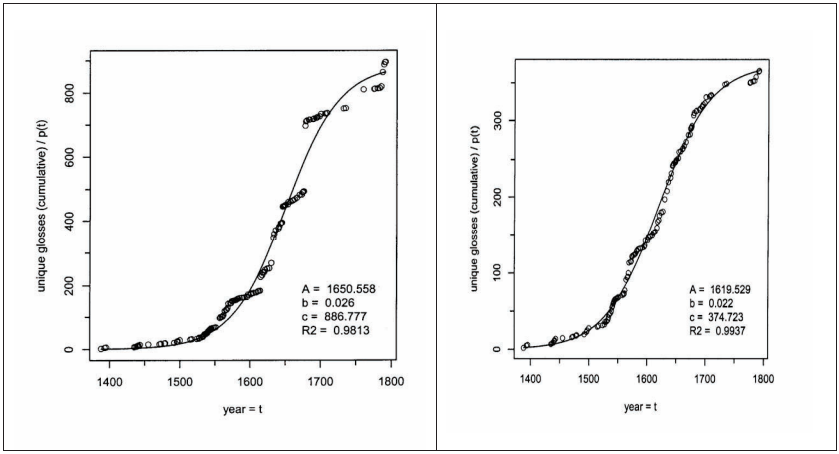
Figure 5. Fitting eq. 3 to grouped Hungarian data

Table 1
Hungarian data

century	*t*	observed	cumulative	predicted
16/1	1	17	17	47.349
16/2	2	39	56	92.851
16/3	3	137	193	175.735
16/4	4	129	322	312.762
17/1	5	198	520	506.654
17/2	6	221	741	728.225
17/3	7	133	874	927.248
17/4	8	227	1101	1070.89
$A = 5.555$ $b = 0.712$ $c = 1258.856$ $R^2 = 0.9939$				

2.2.2 Polish

Figures 6a and 6b show the Polish empirical data as points and the fitted curves (eq. 3, fitted with *R*). The fits are quite good, with $R^2 = 0.9813$ and 0.9937, respectively. The difference between the two sets is negligible, which appears to be a visually convincing confirmation of the validity of the Piotrovskij-Altmann law.



6a. The entire dataset 6b. The trimmed subset

Figure 6. Fitting eq. 3 to the Polish data

As opposed to the Hungarian data in 2.2.1, here there had not been a single, long period of intensified contact within which more than one sigmoid could be found. Rather, there had been a long sequence of consecutive shorter ones. The fractal-like course of the influx is perhaps less striking but visible nonetheless. At least five component sigmoids can be easily discerned and approximated rather accurately with a multi-logistic curve (with $R^2 = 0.9987$ versus 0.9813 with a single one), as illustrated in fig. 7. This is hardly surprising. A much more insightful observation, which, incidentally, would have been lost had the years been grouped, is that the entire dataset can still be modelled with quite high fidelity, by just one single sigmoid.

It might also be remarked that the plots are somewhat reminiscent of the punctuated equilibrium theory (Dixon 1997), only with (considerably) shorter periods of stasis. Possibly, some renewal of interest in the concept, combined perhaps with an effort to formulate for it a more precise definition, is desirable.

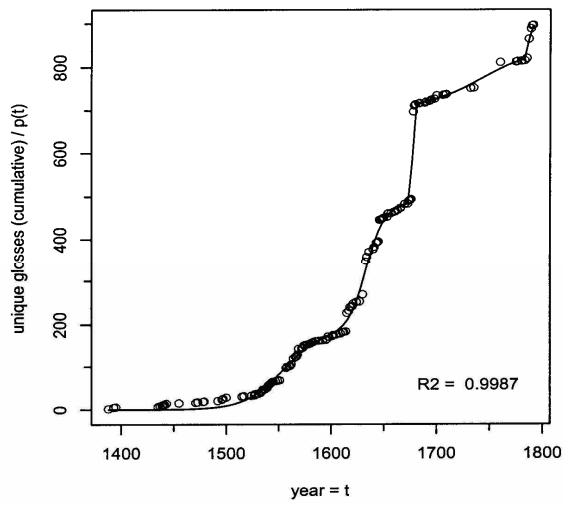


Figure 7. Fitting a multi-logistic curve to the Polish data

Table 2
Coefficients of the five component sigmoids in Fig. 7

$p(t) = \Sigma \frac{c_i}{1 + e^{-b_i(t-A_i)}}$			
i	*Ai*	*bi*	*ci*
1	1555	0.055	185.369
2	1634	0.119	292.535
3	1677	2.820	227.452
4	1742	0.039	135.793
5	1787	2.031	70.472

2.2.3 Conclusions

The three datasets analyzed here can all be approximated very closely by the curve given in equations 2 and 3, which strengthens the empirical support for the

Piotrovskij-Altmann law.

The inconvenient and linguistically meaningless coefficient *a* has been replaced here with an equivalent coefficient $A = ln(a)/b$ which allows the fit to be performed on a timescale in actual years AD, and which is considerably more interpretable from the historical and linguistic point of view.

Also, it has been shown that the Piotrovskij-Altmann law is non-trivial in that it goes in fact against what I believe to be the standard linguistic intuition.

3 Final conclusions

In the present paper, I conducted a qualitative (philological) and a quantitative (Piotrovskij-Altmann law) analysis and attempted to show that the two complement rather than oppose or exclude each other.

It has been suggested (2.1.3) that the coincidence between historical events and linguistic data can probably be considered a correlation, which would enable both linguistics and history to benefit from the results of similar studies. Also, two conjectures have been made on the finer points of this relation.

The two (three) analyzed datasets proved to conform to the Piotrovskij-Altmann law (2.2.3). A slightly modified version of the equation has been proposed, which replaces one of the coefficients with its linguistically and historically more meaningful equivalent. Also, it has been suggested that the purported self-ecidence of the law is, at least to some degree, illusory.

References

Altmann, G. (1983). *Das Piotrovski-Gesetz und seine Verallgemeinerungen.* In: Best/Kohlhase (1983), 59–90.

Best, K.-H. (2003). Spracherwerb, Sprachwandel und Wortschatzwachstum in Texten. Zur Reichweite des Piotrowski-Gesetzes. *Glottometrics 6, 9–34*.

Best, K.-H. (2005). Turzismen im Deutschen. *Glottometrics 11, 56–63*.

Best, K.-H. (2006). Deutsche Entlehnungen im Englischen. *Glottometrics 13, 66–72.*

Best, K.-H. (2008). Sinismen im Deutschen und Englischen. *Glottometrics 17, 87–93.*

Best, K.-H. (2010). Zur Entwicklung des Wortschatzes der deutschen Umgangssprache. *Glottometrics 20, 34–37*.

Best, K.-H., Beöthy E., Altmann G. (1990). Ein methodischer Beitrag zum Piotrowski-Gesetz. *Glottometrika 12, 115–24.*

Best, K.-H., Kohlhase, J. (eds.) (1983). *Exakte Sprachwandelforschung. Theoretische Beiträge, statistische Analysen und Arbeitsberichte* (= *Göttinger Schriften zur Sprach- und Literaturwissenschaft* 2). Göttingen: edition

herodot.

Davies, N. (2005). *God's Playground. A History of Poland*, 1: *The Origins to 1795*, Oxford – New York: Oxford University Press [revised edition].

Dixon, R.M.W. (1997). *The rise and fall of languages*, Cambridge: Cambridge University Press.

Hazai, Gy. (1977). Zur Rolle des Serbischen im Verkehr des Osmanischen Reiches mit Osteuropa im 15.–16. Jahrhundert. In: Décsy Gy., Dimov-Bogoev H.D. (ed.), *Eurasia Nostratica 2, 82–88*, Wiesbaden: Eurolingua.

Hřebíček, L. (1990). *Quantitative studies*. In: Hazai Gy. (ed.), *Handbuch der türkischen Sprachwissenschaft 1, 371–387*. Budapest: Akadémiai Kiadó.

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: Akadémiai Kiadó.

Kálmán, B. (ed.) ([3]1989). *Magyarország történeti kronológiája*, 2: *1526–1848*, Budapest: Akadémiai Kiadó.

Lehfeldt, W., Altmann, G. (2003). Протекание падения редуцированных в древнерусском языке в свете закона Пиотровских. *Russian Linguistics 27, 141–49.*

Leopold, E. (2005). Das Piotrowski-Gesetz (Piotrowski's law). In: Köhler R., Altmann G., Piotrowski R.G. (eds.), *Quantitative Linguistik. Quantitative Linguistics. Ein internationales Handbuch. An International Handbook: 627–633.* Berlin – New York: Walter de Gruyter.

Levyc'kyj, V.V. (2003). Лексична полісемія та квантитативні методи її дослідження. *Мовознавство 4, 17–25*.

Markiewicz, M. (2002). *Historia Polski 1492–1795*. Kraków: Wydawnictwo Literackie.

Molnár, M. (2001). *A Concise History of Hungary.* Cambridge: Cambridge University Press.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*, Vienna: R Foundation for Statistical Computing.

Rycaut/Kłokocki 1678 = **Rycaut P.**, *Monarchia Turecka, Opisana. Przez Ricota Sekretarza Posła Angielskiego. V Porty Ottomanskiey Residviącego. Z Francvskiego Ięzyka Na Polski Przetłvmaczona. Przez Szlachcica Polskiego. Y do Druku podána w Roku 1678.* Słuck: Druk. Radziwiłłowska [transl. H. Kłokocki].

Stachowski, S. (2007). *Słownik historyczny turcyzmów w języku polskim.* Kraków: Księgarnia Akademicka.

Vulanović, R. (2007). Fitting Periphrastic *do* in Affirmative Declaratives. *Journal of Quantitative Linguistics 14(2-3), 111-126.*