Studies in Quantitative Linguistics 23

Issues in Quantitative Linguistics 4

edited by

Emmerich Kelih
Róisín Knight
Ján Mačutek
Andrew Wilson

RAM - Verlag

Issues in Quantitative Linguistics

4

edited by

Emmerich Kelih
Róisín Knight
Ján Mačutek
Andrew Wilson

Dedicated to Reinhard Köhler on the occasion of his 65th birthday

2016

RAM-Verlag


Lecturer
Prof. Dr.

Studies in Quantitative Linguistics

Editors

Fengxiang Fan	(fanfengxiang@yahoo.com)
Emmerich Kelih	(emmerich.kelih@univie.ac.at)
Reinhard Köhler	(koehler@uni-trier.de)
Ján Mačutek	(jmacutek@yahoo.com)
Eric S. Wheeler	(wheeler@ericwheeler.ca)

1. U. Strauss, F. Fan, G. Altmann, *Problems in quantitative linguistics 1*. 2008, VIII + 134 pp.
2. V. Altmann, G. Altmann, *Anleitung zu quantitativen Textanalysen. Methoden und Anwendungen.* 2008, IV+193 pp.
3. I.-I. Popescu, J. Mačutek, G. Altmann, *Aspects of word frequencies*. 2009, IV +198 pp.
4. R. Köhler, G. Altmann, *Problems in quantitative linguistics 2*. 2009, VII + 142 pp.
5. R. Köhler (ed.), *Issues in Quantitative Linguistics*. 2009, VI + 205 pp.
6. A. Tuzzi, I.-I. Popescu, G.Altmann, *Quantitative aspects of Italian texts*. 2010, IV+161 pp.
7. F. Fan, Y. Deng, *Quantitative linguistic computing with Perl.* 2010, VIII + 205 pp.
8. I.-I. Popescu et al., *Vectors and codes of text*. 2010, III + 162 pp.
9. F. Fan, *Data processing and management for quantitative linguistics with Foxpro***.** 2010, V + 233 pp.
10. I.-I. Popescu, R. Čech, G. Altmann, *The lambda-structure of texts*. 2011, II + 181 pp
11. E. Kelih et al. (eds.), *Issues in Quantitative Linguistics Vol. 2*. 2011, IV + 188 pp.
12. R. Čech, G. Altmann, *Problems in quantitative linguistics 3*. 2011, VI + 168 pp.
13. R. Köhler, G. Altmann (eds.), *Issues in Quantitative Linguistics Vol 3*. 2013, IV + 403 pp.
14. R. Köhler, G. Altmann, *Problems in Quantitative Linguistics Vol. 4*. 2014, VIII+148 pp.
15. Best, K.-H., Kelih, E. (eds.), *Entlehnungen und Fremdwörter: Quantitative Aspekte.* 2014. VI + 163 pp.
16. I.-I. Popescu, K.-H. Best, G. Altmann, G. *Unified Modeling of Length in Language.* 2014, VIII + 123 pp.

17. G. Altmann, R. Čech, J. Mačutek, L. Uhlířová (eds.), *Empirical Approaches to Text and Language Analysis dedicated to Luděk Hřebíček on the occasion of his 80th birthday*. 2014. VI + 231 pp.
18. M. Kubát, V. Matlach, R. Čech, *QUITA Quantitative Index Text Analyzer*. 2014, VII + 106 pp.
19. K.-H. Best, *Studien zur Geschichte der Quantitativen Linguistik.* 2015. III + 158 pp.
20. P. Zörnig, K. Stachowski, I.-I. Popescu, T. Mosavi Miangah, P. Mohanty, E. Kelih, R. Chen, G. Altmann, *Descriptiveness, Activity and Nominality in Formalized Text Sequences.* 2015. IV+120 pp.
21. G. Altmann, *Problems in Quantitative Linguistics Vol. 5*. 2015. III+146 pp.
22. P. Zörnig, K. Stachowski, I.-I. Popescu, T. Mosavi Miangah, R. Chen, G. Altmann, *Positional occurrences in texts: Weighted Consensus Strings.* 2015. II+178 pp.

ISBN: 978-3-942303-44-6

RAM-Verlag
Stüttinghauser Ringstr. 44
D-58515 Lüdenscheid
Germany
RAM-Verlag@t-online.de
http://ram-verlag.eu

Contents

The first steps

To learn quantitative linguistics means to collect the works written by Reinhard Köhler and read them thoroughly. Needless to say, one must already know what other (= normal) linguists know, i.e. definitions, classifications, rules, history, etc., but, if one ventures to take a step further, one necessarily bumps against the door of a world one wanted to avoid. It is surely not the infinite paradise but rather the hell in which Köhler, dressed as Lucifer, leads the innocent linguist through the labyrinth of new concepts, formulas, systems, synergetics, statistics, hypotheses, theories. The world of language begins to change its form and the linguist begins (very) slowly to see that there is more light in this world than outside. Köhler loses slowly his Luciferian shape and at once he seems to be the angel responsible for this world. He leads the linguist along ways that join the individual properties of language. One imagines that one is in phonology – very far from syntax – but the angel shows one that there is a very short way between them, and, what is more, one can express it by a formula. That means one need not walk the way; it is sufficient to think it.

Of course, this is rather an esoteric world that exists also in physics and biology, and there is nobody who could prohibit you from entering it (the only exception is the dean of your faculty or head of your department!), but fortunately there is an angel looking sharply at the dean (or head) and leading you into a world in which human language looks like a self-organized system. You may settle on a concept and the system shows you immediately all the links to other concepts. What is more, it brings you to concepts developed by the "Lucifer" himself. As a (pure) linguist, you never heard of them before.

If you succeed in abandoning this world, you will realize that, in front of the door, the same Lucifer stays and wishes you – smiling politely – good moral conscience and much success in repeating what you must teach the students in order to become an "expert" yourself. Linguistics is not about teaching a language but an immersion into worlds which are abstract and similar to the fifth (or higher) dimension of physics. At each step, you can see the smiling Lucifer who created it himself – perhaps only in order to irritate classical linguists who thought that they knew everything already.

But, if you want to stay in this world, you will soon see that you must learn a lot of mathematics, a lot of philosophy of science, forget grammatical rules and all degenerative drawing of trees, and ask the masked Lucifer for help. He will merely smile and, since he wants to retire this year, he will show you a monumental heap of paper and say: "Read all the papers and books in this heap. You can read them more quickly than I wrote them!"

Nevertheless, we hope that he will make the heap higher.

Gabriel Altmann

The Quantitative Nature of *Working Maps* (WM) and *Taxatorial Areas* (TA). A Brief Look at two Basic Units of *Salzburg Dialectometry* (S-DM)

Hans Goebl, Pavel Smečka

1. Preliminary Remarks

Ultimately Salzburg Dialectometry (S-DM) is nothing other than a continuation of classical Romance linguistic geography by quantitative terms and means, obviously assisted by numerical and visual computing (visualistics). In this sense, we would underscore the great epistemic continuity between its theoretical and practical orientations and those of the founding fathers of Romance linguistic geography, such as Jules Gilliéron (1854-1926), the author of the French linguistic atlas ALF (published 1902-1910), and Karl Jaberg (1877-1958) and Jakob Jud (1882-1952), the authors of the Italian linguistic atlas AIS (published 1928-1940).

When speaking about linguistic geography in general, it's necessary to point out a few major methodological differences that characterize this discipline as it is practiced in different Modern Philologies (such as Romance, Germanic, English or Slavonic geolinguistics).

The central methodological key of all varieties of geolinguistics was *first* the study of the diffusion areas of single linguistic traits (features, attributes or characters) and their systematic collection in the form of "linguistic atlases", and *second* the subsequent discovery of their untamable spontaneity (independence or unpredictability) in space, which has given many linguists headaches over the last two centuries because of their chaotic or "irregular" nature. Their claim was therefore often that "dialects" cannot be classified (cf. H. Schuchardt 1870/1900) or simply do not exist (G. Paris 1881). On the contrary, S-DM has from the very beginning accepted fully the "spontaneous" properties of the above mentioned linguistic traits and consequently also the Protean nature of their geographic implementation. The Salzburg term for these allegedly chaotic surfaces is "taxatorial area(s)" (TA).

It should also be said that since the uprise of linguistic geography there have also been several elementary differences between the first linguistic atlases. Whereas Georg Wenker's (1852-1911) monumental "Deutscher Sprachatlas" (DSA) with its more than 50 000 inquiry points tended toward mainly data collection for its own sake, its Gallo-Romance counterpart the ALF originated in the old French tradition of geodetic measurement of the national territory by means of different variables. So, the geolinguistic "measurement" of France (and surrounding zones) - done by Jules Gilliéron and Edmond Edmont between 1897

and 1901 - can rather be qualified as "glotto-geodesy" than a mere collection of dialect data. Obviously this "geometric" underpinning of the ALF is an excellent prerequisite for later dialectometric processing.

One should be aware of the fact that in Gallo-Romance linguistics there was also a great hunger for interesting lexical and phonetic data, but that this hunger has not been stilled by compiling many linguistic atlases but mainly by elaborating on a great number of dialect dictionaries. Immediately after the publication of the ALF, the inner empirical differences between linguistic atlases and dialect vocabularies were discussed thoroughly[1]. It became clear that the proper value of the data of linguistic atlases is rather "relational" than purely documentary, and that they allow for numerous interdisciplinary comparisons, mainly with other space-related sciences[2].

Another peculiarity of Romance linguistic atlases should be underscored: since the publication of the ALF, their data is always exhibited in *full-text maps* and therefore in the original form, thus avoiding any visual or cartographic simplification. As a consequence Romance scholars have been forced, since the dawn of Romance geolinguistics, to use special cartographic techniques in order to get clear the spatial structure of the data laid out on the original atlas maps.

They did it by filling up *silent* (or: *mute*) *maps* of the respective atlas grid (see Jaberg 1906, passim). Obviously, this kind of work requires the training of some very useful linguistic skills such as *simplification* and *classification* of complex geolinguistic data, not to mention the practical challenge of a fairly readable *cartography*. As linguistic geography represented a central discipline in Romance linguistics during the first half of 20th century, many Romance scholars involuntarily became good data classifiers and map makers.

In contrast, Germanic (and other) linguistic atlases never contained their data in their original but instead in graphically encoded form. Their users were therefore not furnished with the same amount of practical and theoretical impulses and stimuli.

2. Taxatation: from the Original Atlas Maps to the Working Maps (WM)

Let us have a look at Map 1 (see Appendix) which is a good example of what has been done by many Romance scholars when classifying and discussing the content of single maps in the ALF (or other linguistic atlases). Even before WW I, Jules Gilliéron published colour maps of the same type (see e. g. Gilliéron/ Mongin 1905). Our map shows the spatial distribution of 15 different denominations of the *ewe* (Fr. *brebis*)[3]. Thus the linguistic nature of this map is *lexical*. The respective 15 diffusion areas vary greatly according to *size*, *shape* (compactness vs. coherence) and *geographic location*. Three major lexical types (in

[1] See Wartburg 1963, 159-163.
[2] See Goebl 2002b and 2006b.
[3] See Wartburg 1918.

Salzburg terminology: *taxates*) emerge: *brebis* (type/taxate 1), *ouaille* (type/ taxate 2) and *fedo* (type/taxate 3). Obviously, they can all be analyzed from an etymological and historical (diachronous) point of view. Etymologically they derive from three well-known Latin roots: VERVÍCE "mutton" > *brebis*, OVÍCULA "(little) sheep" > *ouaille*, and FÉTA "dam" > *fedo*. From a diachronic perspective, the spatial entanglement of the three great areas allows for the hypothetical reconstruction of the geographical spread of the three words over time. These kinds of reconstructive considerations have a very long tradition, not just in the field of Romance linguistics.

In Salzburg, similar colorful maps, provided with a well-defined linguistic interest, are traditionally called *working maps-WM* (Ger. *Arbeitskarten,* Fr. *cartes de travail,* It. *carte di lavoro* etc.). The "work" that has to be done while elaborating such a map relies fully on the theoretical competence of the dialectometrician.

In terms of Numerical Classification, the content of a WM correlates with a single row of the *data matrix*. Metrologically speaking the nature of these data is *qualitative*: they all lie on the *nominal* (or *cardinal*) measurement scale.

As "normal" linguistic atlases consist of several hundreds of *original* maps, each dialectometrization of such an atlas creates at least a similar number of *WM*. Note that the ratio 1:1 between *original map* and *WM* holds for *lexical* maps only. In the case of atlas maps that contain only one *lexical* type and are therefore *mononym*, one can derive several *phonetic* WM (each with different spatial structures) from *one original* atlas map. So the ALF map 233 *chanter* (< Lat. CANTÁRE "to sing") can be split up into several *phonetic* WM, showing respectively the geographic distribution of the Gallo-Romance results of initial C-, pretonic -A, intervocalic -NT-, stressed Á-, intervocalic -R-, and final -E.

Given that the ratio of one taxate/one WM does not create any variation, the taxatorial granulation (or *poly-nymy*) of a WM can theoretically vary between 2 (*bi-nymic* or *2-nymic* WM) and N. The grid of Map 1 (see Appendix) shows 641 (= N) inquiry points, of which 638 belong to the original atlas grid of the ALF, whereas 3 supplementary points (corresponding to the literary languages *French*, *Italian*, and *Catalan*) has been added for illustrative purposes. Incidentally Map 1 is *15-nymic*.

One should be aware of the fact that the whole taxatation process depends on the expert knowledge of the responsible linguist: see Table 1 (below). In the case of the ALF, which is a regular linguistic atlas containing geolinguistic *raw* data in *cartographic* arrangement, this responsibility was exclusively ours; in that of the below mentioned four English "atlases" AES, LAE, CLAE, and WGE the situation is quite more complex. On the one hand, they contain only classified (in Salzburg terminology: taxated) dialect material, and rely, on the other hand, on a source of dialect data ("Survey of English Dialects" – SED) that comprises the original raw data only in *tabular* – and therefore not in *cartographic* – form.

This circumstance complicates the data evaluation enormously, irrespective of the fact that the scientific responsibility is shared by 11 linguists who all pursued diverging scientific interests working independently from each other,

and even at different times (see Table 1, below). Nevertheless, these aggravating prerequisites could not eclipse the inner regularities of the whole data set.

3. The Protean Nature of Taxatorial Areas

Very early (geo)linguists discovered with great amazement the incredible multiformity of TA, even in cases where the categorical proximity of two etymologically related TA suggested a perfect coincidence of the two surfaces and their surrounding lines ("isoglosses"). Such a "categorical proximity" exists, e. g., among the ALF maps 250 *chat* (< *CATTU) "cat", 225 *champ* (< CAMPU) "field", 228 *Chandeleur* (< CANDELÓRU) "Candlemas", 229 *chandelle* (< *CANDÍLLA) "candle", 231 *chanson* (< CANTIÓNE) "song", and 221 *chaîne* (< CATÉNA) "chain", which all show results of initial Latin C+A. Against any (theoretical) expectation, the respective TA (and their surrounding lines) are far from being identical (or prone to coincide)[4].

This fact was first discovered and duly commented in 1889 by Georg Wenker, the author of the German linguistic atlas DSA: "Sind so sämtliche Formen, in denen das Wort erscheint, kartographisch verzeichnet, so werden die einzelnen zu Gruppen sich zusammenschließenden Abweichungen[5] durch Linien abgegrenzt, mit verschiedenen Farben kenntlich gemacht und so das Ganze zu einem übersichtlichen Bilde gestaltet. […] Dann geschieht die Uebertragung in die Grundkarten des Sprachatlas[6], zu denen ein erläuternder Text hinzutritt. Jedes einzelne Wort wird also ganz unabhängig von allen anderen, selbst von verwandten, zu Ende verarbeitet, dann erst werden seine Grenzlinien[7] und seine verschiedenen Formen verglichen mit verwandten Erscheinungen ähnlicher Wörter. Es ist dies eine Vorsicht, welche erst im Verlauf der Arbeit zum Grundsatz erhoben worden ist. *Anfänglich war ich wie wohl jeder allzusehr geneigt, von der bequemen und naheliegenden Vorstellung auszugehen, daß verwandte Wörter, etwa* Hund *und* Pfund*,* Wurst *und* Durst *auch in ihren mundartlichen Eigenheiten zusammenstimmen müßten. Indessen stellte sich heraus, daß dies nicht immer der Fall ist, daß zwar jedes einzelne Wort seine meist ganz festen Grenzlinien besitzt, daß die Grenzlinien verschiedener Wörter dagegen selbst da, wo man es ganz bestimmt erwartet, nicht immer zusammen fallen, sondern bald mehr bald weniger abweichen. Dies allgemeine Ergebniß muß zunächst, gerade wegen seines Gegensatzes zu den bisherigen Anschauungen, nachdrücklich betont werden, bis man sich an diese etwas unbequeme Thatsache gewöhnt hat.* [italics: HG/PS]" (Wenker 2013 [1889], 10).

A similar statement was made by Karl Jaberg in 1908 in his masterly presentation of the ALF: since that time the principle that "each word has its history of its own" (Fr. "chaque mot a son histoire", Ger. "Jedes Wort hat seine

[4] See Jaberg 1908, map III, Wartburg 1963, 22-24, and Berschin/Felixberger/Goebl 2008, 254-256.
[5] In Salzburg terminology: *taxates*.
[6] This base map ("Grundkarte") corresponds to a "silent map" as described above. The final version of the DSA base map had a grid with more than 50 000 inquiry points.
[7] I. e. *isoglosses*.

eigene Geschichte") has reigned, which in our case should be adapted in "each *taxate* has its *area* of its own"[8]. Starting from this principle, many linguists conjectured – unfortunately – that behind many TA there is absolutely no order or regularity whatsoever. We will see that this belief was pure superstition.

In summary it can be stated that

- TA vary enormously along *size*, *shape* and *location*.
- on mute maps (with N polygons) the *size* of TA oscillates between 1 and N-1.
- the quantitative measurement of their *shape* and *location* seems to be currently out of range.

4. The "Special Entanglement" of Taxatorial Areas

Obviously, the continuous change in *size*, *shape* and *location* of the different TA goes back to a wide range of intra- and extra-linguistic causes, some of which can be detected and even "explained". But what should now be clear is the fact that this variegated situation is not "unnatural", nor is it the consequence of a series of catastrophes that destroyed an assumed virgin regularity. Superposing a greater series of TA and controlling their spatial deviations from each other one discovers that all these TA are interlocked together like shingles on a roof[9]. As we found this phenomenon in all our dialectometric analyses, it seems appropriate to denominate it properly: we proposed for it the following terms: *special entanglement*, Ger. *spezielle Verzahnung*, Fr. *enchevêtrement particulier*, It. *intreccio particolare*, Sp. *entramado especial* etc.

It's highly probable that the special entanglement that also occurs in many other geo-based sciences belongs to the founding principles of all spatial networks. It's not less probable that it is a direct consequence of diversification processes that operate continuously in such reticulated structures. G. Altmann (1985) modeled these circumstances, referring to the two main Zipfian processes of *diversification* and *unification*, which permanently alter the inner structure of geolinguistic grids[10]. In particular, Altmann conjectured the interplay of *birth*-and-*death*-processes that created or annihilated the single components (or *taxates*) of the respective network.

From the standpoint of sociology and sociolinguistics the phenomenon of special entanglement can be interpreted as a consequence of a particular communicative behavior of man – generally conceived as HOMO *LOQUENS* – in space. For many years now[11] we've called it the "basilectal management of space by man". Note that similar concepts also exist in geography and anthropology

[8] For a thorough discussion of this question and the related problems see Malkiel 1967 and Christmann 1971.

[9] The same phenomenon was addressed in 1876 by the Italian linguist G. I. Ascoli (1829-1907) when he claimed that « dialects », conceived as geotypological concepts, are defined by a « particolar combinazione » of a set of linguistic traits.

[10] See also the bibliography of diversification compiled by K.-H. Best (2014).

[11] Cf. Goebl 1993, 277.

where the idea that the natural dimensions of space can be altered by human activity is very common.

A	B	C	D	E	F	G	H	I
Linguistic Atlas	**Number of Original Maps**	**Number of analyzed Original Maps**	**Number of classifiers (England : atlas authors)**	**Number of analyzed Working Maps**	**Number of Inquiry Points**	**Number of Taxates/ Taxatorial Areas**	**Average Size of the Taxatorial Areas (= E×F/G)**	**Range of Polynymy from 2 to *x***
ALF	1 421	626	1 (HG)	1 681	641	19 328	55,74	90
England TOT	1 711	1 516	11	1 524	313	16 810	28,37	108
AES	424	424	4	424	313	5 838	22,73	39
CLAE (I+II)	315×2 (2 taxatorial levels)	591	2	597	313+1	7 698	24,35	108
LAE	406	388	3	389	313+1	2 839	43,02	21
WGE	251	114	2	114	313+1	435	82,28	7

Table 1. Empirical and taxatorial characteristics of one French (ALF) and four English linguistic atlases (AES, CLAE [I and II], LAE, and WGE).

5. Evidence from French Dialects (ALF)

See Table 1 (above) and Appendix (Map 1, Figure 1 and Figure 2).

We will now show two law-like regularities, in the stock of WM and TA in the ALF that occurred in very similar form in all our dialectometric analyses of a great number of Romance, Germanic and English linguistic atlases.

The 1 421 maps of the series A of the ALF contain 638 inquiry points and show the results of 639 inquiries done by the fieldworker Edmond Edmont (1849-1926) between 1897 and 1901[12]. They cover all linguistic categories, from phonetics to syntax. Between 1996 and 1999, 626 out of these 1 421 maps were analyzed in Salzburg for dialectometric purposes[13]. The result is 1 681 WM containing 19 328 taxates and their respective areas. They still cover all linguistic categories.

Regarding the 1 681 WM, Figure 1 shows the very regular relationship between the increase in their inner polynymy, and the decrease in their absolute frequency. In other words, many WM have very simple structures; very few WM offer highly variegated structures.

[12] For the ALF in general see Brun-Trigaud/Le Berre/Le Dû 2005.

[13] The results of the dialectometrization of the ALF have been presented in a long series of articles published in different languages : see our contributions from 2000 to 2014b.

This relation was studied by G. Altmann in 1985 under the assumption of the permanent pull of self-regulating birth-death-processes. As a result, he defined the so-called "Goebl-Law", which applies to diversification-processes in geolinguistic data[14].

From a merely geolinguistic point of view, Figure 1 very clearly shows the percentages of "beautiful" and "chaotic" maps in an atlas corpus. Normally, geolinguistic handbooks and readers discuss only "beautiful" maps (with reduced polynymy between 2 to 10), neglecting completely highly polynymic maps for reasons of excessive "chaoticity". Nevertheless, it was shown with dialectometric means that corpuses with, one the one hand, *low*-polynymic and, on the other, *high*-polynymic WM contain exactly the same deep structures[15]. This proves that there is a great amount of *redundancy* in the global geolinguistic deep structures.

The same effect was demonstrated in the 1980s by manipulating systematically the *quantity* of WM to be combined dialectometrically. It was shown that the overall deep structures of the respective data stocks already appeared from the synthesis of approx. 200 randomly chosen WM[16].

Figure 2 shows the frequency distribution of 19 328 taxatorial areas whose geographic *size* oscillates between 1 (with high frequency) and 640 (= N – 1) (with low frequency). Our assumption is that the above mentioned Zipfian forces are also responsible in this case for the apparent regularity of the curve.

We remember that the aforementioned ALF data were elaborated in a special research project realized under the exclusive responsibility of H. Goebl. So, the collected data reflect what could be called his personal "geolinguistic bias". In the next chapter it will be shown that the same results emerge when combining different "personal geolinguistic biases".

6. Evidence from English Dialects (AES, CLAE, LAE, WGE)

See Table 1 (above) and Appendix (Map 1, Figure 3 and Figure 4).

The history of English geolinguistics is completely independent from French geolinguistics, both its practical experiences and its brilliant scientific achievements. The respective fieldwork began in France in 1897 and did not commence in England until 1950. It's very strange to see that the younger English initiative neglected completely the lessons of the older French one. The name of the English initiative is "Survey of English dialects" (SED). One of the strangest peculiarities of the English initiative was the publication of the collected data (embracing only 313 inquiry points) not in *cartographic* but exclusively in *tabular* form. So, the 12 data volumes of SED, published between 1962 and 1971, are far from being as suggestive as the large in folio maps of the ALF or

[14] In his study Altmann referred to analyses presented in Goebl 1984, based themselves on data drawn from the linguistic atlases AIS and ALF.
[15] See Goebl 2014a.
[16] See Goebl 1984 I, 206 ss.

AIS. So, English linguists could not benefit from the illuminating effects of *full-text maps* and their current (generalized) elaboration by means of *mute maps*.

As a consequence some English linguists became "privileged" interpreters of selected portions of the SED data, by publishing the results of their classificatory analyses of a certain amount of SED tables under the slightly misleading title "Linguistic atlas of…". This procedure holds for the WGE (1974), LAE (1978), AES (1979), and CLAE (1991, 1997) "atlases", none of which contains original dialect data, but instead coded maps.

CLAE had the benefit of being produced already by electronic means. Its author, Wolfgang Viereck (Bamberg), handed us over the electronic files of the two volumes of CLAE in the 1990s for further dialectometric analyses[17]. Given the particularly interesting results of the dialectometrization of these data, we subsequently decided to grasp the data of other similar English "linguistic atlases", all derived from SED. Although the respective data entry was rather laborious, it was well accomplished thanks to the precision and energy of our Salzburg collaborators.

It should be emphasized that the data collection generated in this way reflects the classificatory "philosophy" of 11 different Anglicists, embracing all linguistic categories.

Nevertheless, Figures 3 and 4 show exactly the same quantitative tendencies that we already saw in case of the ALF. In Figure 3 the polynymy of the 1 524 WM goes from 2 to 108[18], whereas in Figure 4 the size of the 16 810 TA varies between 1 and 310 (inquiry points or polygons). Obviously, the numbers shown in Table 1 (see above) for "England TOT" represent the sum of those of the AES, CLAE, LAE, and WGE.

Note that the 315 coded maps of the two volumes of the CLAE show two taxatation levels: a "lumped" (i.e. with a more coarse structure) and a "split" (i.e. with finer granulation) one. Thus, the number of the respective WM has been doubled.

In a nutshell, it seems to be evident that simply counting the frequency of the two basic units of dialectometric data matrices – *working maps* (WM) and *taxates* or *taxatorial areas* (TA) – produces clear-cut quantitative regularities that reflect some elementary properties of the dialectal behavior of man in space.

7. Concluding Remarks

Summing up, we like to emphasize some historical facts. One of the greatest discoveries of the last quarter of the 19th century was the theory of the general regularity of sound change. This remains true despite the many exaggerations and confusions that have been perpetrated since then. The central point of these discussions was the analysis of the change of linguistic utterances along the axis of *time*, done under the tacit assumption that *time* represents an absolute term.

[17] See Goebl 1997 and Goebl/Schiltz 1997.

[18] The polynymy 108 occurs on the map S9 ("I know a man [*who*] will do it for you.") of CLAE I (of 1991).

One of the errors committed during this time was the claim that a particular sound change in the language x must be valid also in all geographical varieties (dialects) of the same language. Many maps of the early linguistic atlases proved that this argument was wrong.

Even though the faultiness of this argument has been proved time and again, no serious discussions arose on the relationship between linguistic behavior in *time* and *space* and to what extent the famous linearity of *time* and the orderly structure of time-related linguistic utterances could have a counterpart in *space*. Obviously, coping with the challenges of *time* was much easier than coping with those of *space*. So almost hundred years passed between the beginning of the sound law discussions in the circle of the Leipzig neogrammarians (1876)[19] and the earliest publication of genuine dialectometry (Séguy 1971).

Nowadays, it should be taken for granted that a language evolves in *space* under the same constraints of "non-chaoticity" as it does in *time* (and perhaps also in other dimensions).

8. Acknowledgments

Drawing of Map 1: Yves Scherrer, Geneva,
Drawing of the Figures 1-4: Werner Goebl, Vienna,
Stylistic supervision: Benjamin Wright, Salzburg.

References

AES: Kolb, E., Glauser, B., Elmer, W., and Stamm, R. (eds.) (1979). *Atlas of English Sounds*, Bern: Francke.

AIS: Jaberg, K., Jud, J. (eds.) (1928-1940). *Sprach- und Sachatlas Italiens und der Südschweiz*, 8 vols., Zofingen: Ringier (reprint: Nendeln: Kraus, 1971) (web-based version: http://www3.pd.istc.cnr.it/navigais/).

ALF: Gilliéron, J., Edmont, E. (eds.) (1902-1910). *Atlas linguistique de la France*, 10 vols., Paris: Champion (reprint: Bologna: Forni, 1968) (two web-based versions: 1 : http://diglib.uibk.ac.at/urn:nbn:at:at-ubi:2-4568; 2 : http://cartodialect.imag.fr/cartoDialect/carteTheme).

Altmann, G. (1985). Die Entstehung diatopischer Varianten. *Zeitschrift für Sprachwissenschaft* 4, 139-155.

Ascoli, G. I. (1876). Paul Meyer e il franco-provenzale. *Archivio glottologico italiano* 2, 385-395.

Berschin, H., Felixberger, J., Goebl, H. (2008[2]). *Französische Sprachgeschichte. Lateinische Basis, interne und externe Geschichte, sprachliche Gliederung Frankreichs. Mit einer Einführung in die historische Sprachwissenschaft*, Hildesheim: Olms.

[19] This year saw the appearance of the famous treatise of August Leskien on the "Deklination im Slawisch-Litauischen und Germanischen" among others; for further details see Putschke 2001, 1508.

Best, K.-H. (2014). Diversification: Bibliography. *Glottometrics* 28, 87-91.

Brun-Trigaud, G., Le Berre, Y., Le Dû, J. (2005). *Lectures de l'Atlas linguistique de la France de Gilliéron et Edmont. Du temps dans l'espace. Essai d'interprétation des cartes de l'*Atlas linguistique de la France *de Jules Gilliéron et Edmond Edmont augmenté de quelques cartes de l'*Atlas linguistique de la Basse-Bretagne *de Pierre Le Roux*, Paris: CTHS.

CLAE: Viereck, W., Ramisch, H. (eds.) (1991, 1997). *The Computer Developed Linguistic Atlas of England*, Tübingen: Niemeyer, 2 vols.

Christmann, H. H. (1971). Lautgesetze und Wortgeschichte. Zu dem Satz „Jedes Wort hat seine eigene Geschichte". In: *Sprache und Geschichte. Festschrift für Harri Meier zum 65. Geburtstag*, Coseriu, E., Stempel, W.-D. (eds.), München: Fink, 111-124.

DSA: *Deutscher Sprachatlas, aufgrund des von Georg Wenker begründeten Sprachatlas des Deutschen Reiches in vereinfachter Form begonnen von Ferdinand Wrede, fortgesetzt von Walther Mitzka und Bernhard Martin,* Marburg/Lahn: Elwert 1927-1956, 4 vols. (23 fascicles with 128 maps).

Gilliéron, J., Mongin, J. (1905). *Scier dans la Gaule romane du sud et de l'est. Étude de géographie linguistique*, Paris: Champion (Italian translation: *« Segare » nella Gallia romanza meridionale e orientale*, by Lorenzo Massobrio, Novi Ligure: Grafica editoriale universitararia, 1990).

Goebl, H. (1984). *Dialektometrische Studien. Anhand italoromanischer, räto-romanischer und galloromanischer Sprachmaterialien aus AIS und ALF*, 3 vols., Tübingen: Niemeyer.

Goebl, H. (1993). Dialectometry. A Short Overview of the Principles and Practice of Quantitative Classification of Linguistic Atlas Data. In: Köhler, R., Rieger, B. B. (eds.). *Contributions to Quantitative Linguistics*, Dordrecht: Kluwer, 277-315.

Goebl, H. (1997). Some Dendrographic Classifications of the Data of CLAE 1 and CLAE 2. In: *CLAE* II, 23-32.

Goebl, H. (2000). La dialectométrisation de l'ALF: présentation des premiers résultats. *Linguistica* 40, 209-236.

Goebl, H. (2002a). Analyse dialectométrique des structures de profondeur de l'ALF. *Revue de Linguistique Romane* 66, 5-63.

Goebl, H. (2002b). Sprachatlanten: woher? womit? wozu? Einige buntgemischte Reflexionen am Gartenzaun zwischen Romanistik und Germanistik. In: *Namen, Sprachen und Kulturen. Imena, Jeziki in Kulture. Festschrift für Heinz Dieter Pohl zum 60. Geburtstag*, Anreiter, P., Ernst, P., Hausner, I., and Kalb, H. (eds.), Wien: Edition Präsens, 257-274.

Goebl, H. (2003). Regards dialectométriques sur les données de l'Atlas linguistique de la France (ALF): relations quantitatives et structures de profondeur. *Estudis Romànics* XXV, 59-121.

Goebl, H. (2005). La dialectométrie corrélative. Un nouvel outil pour l'étude de l'aménagement dialectal de l'espace par l'homme. *Revue de Linguistique Romane* 69, 321-367.

Goebl, H. (2006a). Recent Advances in Salzburg Dialectometry. *Literary and Linguistic Computing* 21/4, 411-435.

Goebl, H. (2006b). Warum die Dialektometrie nur in einem roman(ist)ischen Forschungskontext entstehen konnte. In: Dahmen, W., Holtus, G., Kramer, J., Metzeltin, M., Schweickard, W., and Winkelmann, O. (eds.), *Was kann eine vergleichende romanische Sprachwissenschaft heute (noch) leisten?* Romanistisches Kolloquium XX, Tübingen: Narr, 291-317.

Goebl, H. (2007a). A Bunch of Dialectometric Flowers: a brief Introduction to Dialectometry. In*:* Smit, U., Dollinger, St., Hüttner, J., Kaltenböck, G., and Lutzky, U. (eds.), *Tracing English through Time. Explorations in Language Variation.* In Honour of Herbert Schendl on the Occasion of his 65th Birthday, Wien: Braumüller, 133-171.

Goebl, H. (2007b). Dialectometry: Theoretical Prerequisites, Practical Problems, and Concrete Applications (mainly with Examples drawn from the "Atlas linguistique de la France", 1902-1010). In*: Geolinguistics around the World.* Proceedings of the 14th NIJLA [National Institute of Japanese Language] International Symposium (Tokyo, August 22-23, 2007), Tokyo: NIJLA, 65-74.

Goebl, H. (2009). Quelques coups d'oeil dialectométriques sur l'Atlas linguistique de la France: structures de surface et structures de profondeur. In*:* Dalbera-Stefanaggi, M.-J., Simoni-Aurembou, M.-R. (eds.), *Images de la langue: représentations spatiales, sémantiques et graphiques*, Paris: Editions du CTHS, 39-60.

Goebl, H. (2010). Dialectometry and quantitative mapping. In: Lameli, A., Kehrein, R., and Rabanus, St. (eds.), *Language and Space. An International Handbook of Linguistic Variation*, vol. 2: *Language Mapping* (Handbücher der Sprach- und Kommunikationswissenschaft [HSK] 30.2.), Berlin: de Gruyter; 1st part: 433-457 (text), 2d part (maps): 2201-2212.

Goebl, H. (2012). Introduction aux problèmes et méthodes de l'»École dialectométrique de Salzbourg« (avec des exemples gallo-, italo- et ibéroromans). In: Álvarez Pérez, A., Carrilho, E., and Magro, C. (eds.), *Proceedings of the International Symposium on Limits and Areas in Dialectology (LimiAr)*, Lisbon 2011. www.http://limiar.clul.ul.pt", Lisboa: Centro de Linguística da Universidade de Lisboa, 117-166.

Goebl, H. (2014). L'impact de la polynymie des cartes d'atlas sur le résultat de calculs dialectométriques. In: *Linguistique romane et Linguistique indo-européenne. Mélanges offerts à Witold Mańczak à l'occasion de son 90e anniversaire*, Polska Akademia Umiejetnosci. Instytut Filologii Romanskiej Uniwersytetu Jagiellonskiego (ed.), Kraków: Polska Akademia Umiejetnosci. Instytut Filologii Romanskiej Uniwersytetu Jagiellonskiego, 243-260.

Goebl, H., Schiltz, G. (1997). Dialectometrical Compilation of CLAE 1 and CLAE 2. Isoglosses and Dialect Integration. In: *CLAE* II, 13-21.

Goebl, H., Smečka, P. (2014). L'analyse dialectométrique des cartes de la série B de l'ALF. *Revue de Linguistique Romane* 78, 439-497.

Jaberg, K. (1906). Zum Atlas linguistique de la France. *Zeitschrift für romanische Philologie* 30, 512.

Jaberg, K. (1908). *Sprachgeographie. Beitrag zum Verständnis des Atlas linguistique de la France*, Aarau: Sauerländer (Spanish translation: *Geografía lingüística. Ensayo de interpretación del „Atlas lingüístico de Francia",* by Llorente, A., Alvar, M., Granada: Universidad de Granada. Secretariado de Publicaciones, 1959).

LAE: Orton, H., Sanderson, St., and Widdowson, J. (eds.) (1978). *The Linguistic Atlas of England*, London: Croom Helm.

Malkiel, Y. (1967). Each Word has a History of its Own. *Glossa* 1, 137-149.

Paris, G. (1888). Les parlers de France. In: Paris, G., *Mélanges linguistiques. Latin vulgaire, langues romanes, langue française, notes étymologiques*, publiés par Mario Roques, Paris: Champion, 1909, 432-448.

Putschke, W. (2001). Die Dialektologie, ihr Beitrag zur historischen Sprachwissenschaft im 19. Jahrhundert und ihre Kritik am junggrammatischen Programm. In: Auroux, S., Koerner, E.F.K., Niederehe, H.-J., and Versteegh, K. (eds.), *History of Language Sciences. Geschichte der Sprachwissenschaften. Histoire des sciences du langage. An International Handbook on the Evolution of the Study of Language from the Beginnings to the Present. Ein internationales Handbuch zur Entwicklung der Sprachforschung von den Anfängen bis zur Gegenwart. Manuel international sur l'évolution de l'étude du langage des origines à nos jours*, Berlin, New York: Walter de Gruyter, vol. 2, 1498-1513.

Schuchardt, H. (1870). *Über die Klassifikation der romanischen Mundarten* [Probe-Vorlesung gehalten zu Leipzig am 30. April 1870], Graz: Styria 1900 (also published in: Spitzer, L. (1928[2]). *Hugo Schuchardt-Brevier. Ein Vademecum der allgemeinen Sprachwissenschaft*, Halle: Niemeyer, 166-188; reprint: Tübingen: Niemeyer, and Darmstadt: Wissenschaftliche Buchgesellschaft, 1976).

SED: Orton, H., Halliday, W. J., Dieth, E., and Wakelin, M. F. (eds.) (1962-1971). *Survey of English Dialects*. *The Basic Material*, Leeds: E. J. Arnold, 12 vols. (reprint: London: Routledge, 1998).

Séguy, J. (1971). La relation entre la distance spatiale et la distance lexicale. *Revue de Linguistique Romane* 35, 335-357.

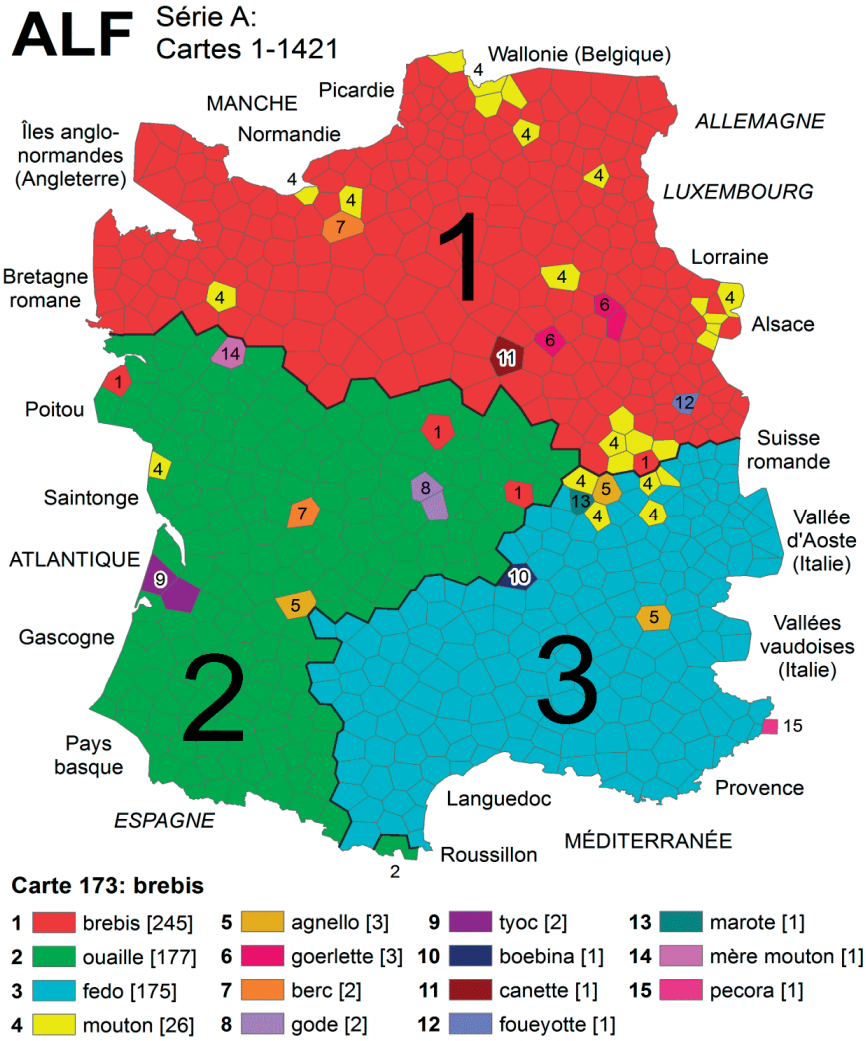
Wartburg, W. v. (1918). *Zur Benennung des Schafes in den romanischen* Sprachen : *ein Beitrag zur Frage der provinziellen Differenzierung des späten Lateins*, Berlin: Reimer.

Wartburg, W. v. (1963[2]). *Problèmes et méthodes de la linguistique*, Paris: Presses Universitaires de France, translated from German by Pierre Maillard with the participation of Stephen Ullmann, second edition.

Wenker, G. (2013). *Schriften zum Sprachatlas des Deutschen Reichs. Gesamtausgabe. Band I: Handschriften: Allgemeine Texte. Kartenkommentare 1889-1897*, published by Alfred Lameli, Hildesheim, Zürich, and New York: Olms.

WGE: Orton, H., Wright, N. (eds.) (1974). *A Word Geography of England*, London, New York: Seminar Press.

Appendix



Map 1: Taxatorial analysis ("working map") of map 173 (*la brebis*) of the ALF showing the geographical distribution of fifteen Gallo-Romance denominations ("geo-synonyms") of the "ewe".

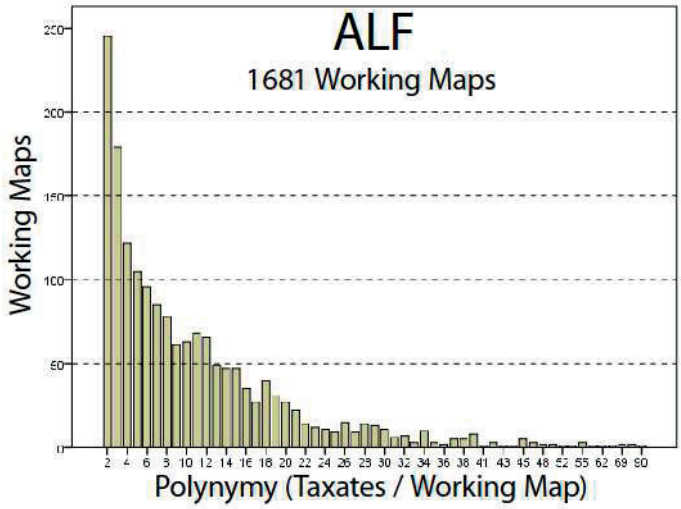


Figure 1. Histogram showing the relationship between geolinguistic polynymy and the number of working maps (WM). Data: 626 original maps of the ALF (1902-1910), taxatation (typification) encompassing all linguistic categories, 1 681 WM. The polynymy oscillates between 2 and 90 taxates per WM; the number of WM varies between 245 (2-nym WM) and 1 (90-nym WM).

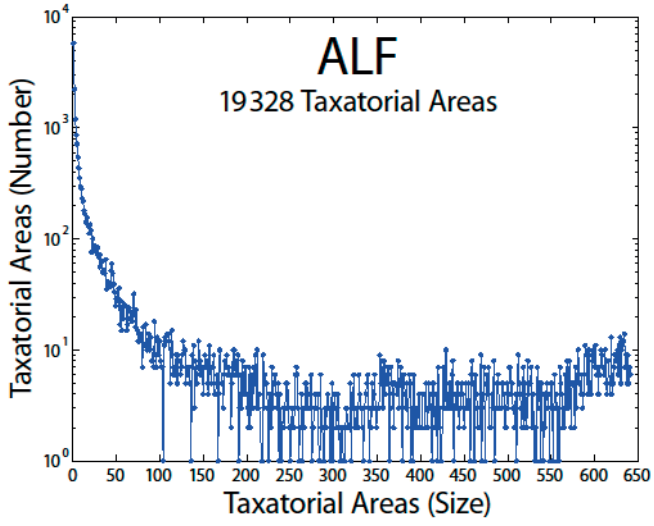


Figure 2. Diagram showing the relationship between size and number of 19 328 taxatorial areas (TA). Data: 626 original maps of the ALF (1902-1910), taxatation (typification) encompassing all linguistic categories, providing 1 681 WM, and 19 328 TA. The size of TA oscillates between 640 (inquiry points or polygons) and 1, their number between 5 743 and 1.

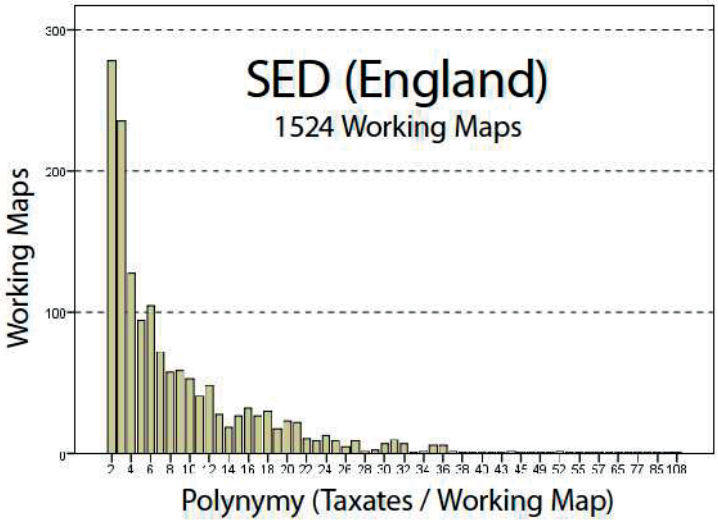


Figure 3. Histogram showing the relationship between geolinguistic polynymy and the number of working maps (WM). Data: 1 516 original maps of the AES, CLAE (I and II), LAE, and WGE, taxatation (typification) encompassing all linguistic categories, providing 1 524 WM. The polynymy oscillates between 2 and 108 taxates per WM; the number of WM varies between 278 and 1.

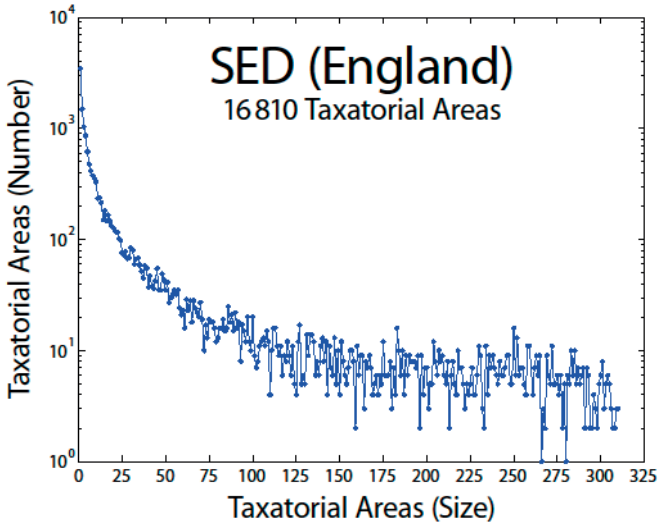


Figure 4: Diagram showing the relationship between the size and the number of 16 810 taxatorial areas (TA). Data: 1 516 original maps of AES, CLAE (I and II), LAE, and WGE, taxatation (typification) encompassing all linguistic categories, providing 1 524 WM, and 16 810 TA. The size of TA oscillates between 310 (inquiry points or polygons) and 1, their number between 3 477 and 1.

ISBN 978-3-942303-44-6