

Sebastian Kempgen

**WORTWOLKEN ALS HEURISTISCHES MITTEL DER
VISUALISIERUNG IN DER SLAVISTISCHEN LINGUISTIK
(AM BEISPIEL VON 40 JAHREN „KONSTANZER KREIS“)**

1. Vorbemerkung

Gegenstand des vorliegenden Beitrages ist zweierlei: im weiteren Sinne die Visualisierung als neue Darstellungstechnik im Kontext der Digital Humanities, und zwar speziell die Nutzung von sog. „Wortwolken“, zweitens ein analytischer Rückblick auf 40 Jahre Vortrags- und Publikationstätigkeit des „Konstanzer Kreises“¹. Diese beiden Punkte sollen miteinander verknüpft werden, d.h. Themen und Entwicklungen der Aktivitäten des Konstanzer Kreises sollen mit den zuvor genannten Verfahren der Visualisierung präsentiert werden.

Mit dem ersten Thema will der Artikel einen Beitrag zur weiteren Entwicklung der Digital Humanities in der Slavistik liefern, für die es natürlich schon vielerlei Beispiele und Belege gegeben hat, bevor der moderne Sammelbegriff für entsprechende Verfahren und Fragestellungen populär wurde, mit dem zweiten soll eine kritische Reflexion der Position des Konstanzer Kreises innerhalb der deutschsprachigen slavistischen Linguistik geleistet werden, nicht etwa ein Eigenlob. Die vorliegende kleine Bilanz der Themen, Aktivitäten und Wirkungen des Konstanzer Kreises ist eine zweite solche Analyse nach derjenigen, die zum 25-jährigen Bestehen von Renate Rathmayr erschien und den Titel trug „«Mir scheint, daß es im Augenblick mehr Fragen als Antworten gibt.» Eine quantitative Analyse des wissenschaftlichen Schreibens anhand von 279 Schlußsätzen und 25 Jahren Konstanzer Kreis“ (Rathmayr 2000 als Publikation zum Vortrag von 1999).

¹ Vgl. den Wikipedia Eintrag <http://de.wikipedia.org/wiki/Konstanzer_Kreis> sowie die Materialien auf <http://kodeks.uni-bamberg.de/Konstanzer_Kreis/Index.html>, wo vom Autor jeweils Vortragsprogramm und bibliographische sowie biographische Angaben archiviert werden.

2. Visualisierung in der (Slavistischen) Linguistik

Visualisierung hat es in der Slavistischen Linguistik seit jeher genau so gegeben wie in der Sprachwissenschaft allgemein. Die klassischen und jedermann gut bekannten graphischen Darstellungen sind z.B. Dialektkarten zum slavischen Sprachraum oder seinen Teilbereichen, Baumdiagramme zur Darstellung der slavischen Sprachfamilie und ihrer genetischen Gliederung und phonetische Sachverhalte. Auch die „Urheimat“ der Slaven und ihre Wanderungsbewegungen haben naheliegenderweise eine Tradition der Visualisierung. Wann und mit welchen Darstellungen aber eigentlich die Visualisierung in der Slavischen Sprachwissenschaft eingesetzt hat, wäre eine interessante wissenschaftsgeschichtliche Fragestellung. Für einen Bereich dürfte dies im übrigen relativ klar sein: August Schleichers bekannter Stammbaum der indogermanischen Sprachen von 1861 ist auf jeden Fall ein Meilenstein in der wissenschaftlichen Visualisierung von Sprachverwandtschaft, denn hier sollen – im Gegensatz zu einigen früheren Darstellungen, die das Bild von Händen oder Bäumen eher metaphernartig nutzten – der Abbildung ja definierte Achsen zugrundeliegen: „Die Länge der Linien deutet die Zeitdauer an, die Entfernung derselben von einander den Verwandtschaftsgrad.“ (Schleicher 1861, 7 in originaler Orthographie; Abb. 1 siehe unten).

Nebenbei: Daß linguistische Visualisierungen eine politische Dimension haben oder bekommen können, ist unschwer auch an einem gegenwärtigen, ganz aktuellen Beispiel zu erkennen, nämlich an Dialektkarten des Ukrainischen. Diese beziehen bislang die Krim selbstverständlich in das Verbreitungsgebiet des Ukrainischen mit ein, auch wenn historische Karten unbestritten ein Überwiegen des Russischen zeigen (etwa im Verhältnis 90% : 10%). Wie lange aber wird es dauern, bis russische Sprachkarten nach der Annexion der Krim revidiert werden – und wird das Ukrainische dort dann noch als Minderheitensprache auftauchen oder nicht? Das wird spannend zu beobachten und das ideologisch diktierte Ergebnis wohl vorhersehbar sein.

Auch quantitative Arbeiten in der Sprachwissenschaft haben immer schon mit Visualisierungen gearbeitet: Häufigkeitsverteilungen, Clusteranalysen, Verläufe usw. lassen sich nachweislich leichter erfassen, wenn eine Kurve oder Graphen die Daten anzeigen, als wenn eine Tabelle die gleichen Informationen enthält. Dies gilt selbstredend auch für die slavische Sprachwissenschaft.

Im übrigen sei daran erinnert, daß Schriftsprache selbst ein Mittel der Visualisierung ist, ursprünglich sogar ganz im Wortsinne, wenn Silbenzeichen (und dann Buchstaben) aus Bildern hergeleitet werden. Eines der rätselhaftesten Beispiele solcher Visualisierung von Sachverhalten ist dabei sicher der Diskus von Phaistos. Ähnlich alte und bedeutende Beispiele gibt es aus dem slavischen Raum zwar nicht, aber auch hier könnte man immerhin doch an gewisse Diskus-

sionen um den Ursprung der Glagolica erinnern, bei der mit den offensichtlichen Komponenten von Dreiecken, Kreisen und Kreuzen ja auch ein symbolischer Gehalt in Rede stand.

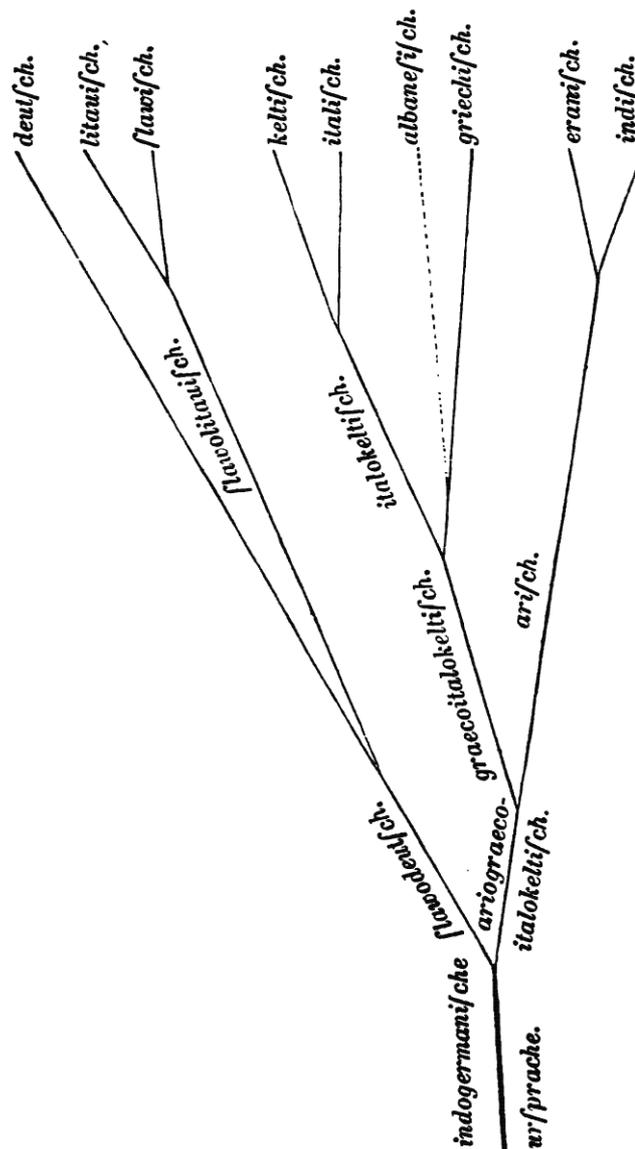


Abb. 1: Frühe Visualisierung: August Schleichers Stammbaum (1861, 7):
 „Die länge der linien deutet die zeitdauer an,
 die entfernung derselben von einander den verwantschaftsgrad.“

3. Digitale Geisteswissenschaften und Visualisierung

Daß sich gegenwärtig eine neue Sub-Disziplin der Sprachwissenschaft zu entwickeln scheint, läßt sich an mehreren Indikatoren ablesen, z.B. daran, daß 2014 eine programmatische Konferenz unter dem Titel „Visuelle Linguistik: Theorie und Anwendung von Visualisierungen in der Sprachwissenschaft“ stattgefunden und unter jungen Linguisten großen Zuspruch gefunden hat.² Allerdings war die Slavistik, soweit zu sehen, hier nicht vertreten. Die Beiträge der Konferenz behandelten Visualisierung sehr oft im Kontext von Korpusanalysen, die ja meist mit großen Datenmengen zu tun haben – und damit ist schon ein wichtiger Anlaß benannt, in dem solche heuristischen Auswertungsverfahren mit Gewinn zum Einsatz kommen können. Die Visualisierung ist damit eine folgerichtige Weiterentwicklung der intensiven Beschäftigung mit umfangreichen Korpora, die es im Westen wie in den slavischen Ländern gegeben hat, wo in der Folge überall Nationalkorpora entstanden sind.

Neben solchen Konferenzen, wie sie jetzt auch in der Slavistik stattfinden,³ sind andere Dinge ein klares Anzeichen dafür, daß in der entsprechenden Wissenschaft eine Entwicklung stattfindet. Dazu gehören etwa programmatische Sammelbände und Monographien, die genau so heißen wie die Disziplin („Digital Humanities“), Lehrstühle und Webseiten, Broschüren, die eine Bestandsaufnahme versuchen – im vorliegenden Fall im deutschen Raum z.B. eine solche, die 2011 an der Universität Köln entstanden und unter dem Titel „Digitale Geisteswissenschaften“ online zu finden ist.⁴ Kennzeichnend ist es auch, wenn große Wissenschaftsorganisationen wie die Alexander-von-Humboldt-Stiftung ein solches Thema auf das Cover ihres Magazins heben (so für das Heft 102/2014 von „Humboldtkosmos“).

Neu sind auch Formen der Visualisierung, die in jüngster Zeit gerade im Wissenschaftsjournalismus einem allgemeinen Leserkreis präsentiert werden. Meist handelt es sich dabei um die Darstellung von Netzwerken und Zusam-

² Vgl. <http://www.visual-linguistics.net/symposium/index_en.html> mit weiteren Informationen. Wenn der *Call for Papers* zu dieser Konferenz allerdings die „lange Tradition“ der Visualisierung in der Sprachwissenschaft mit dem Hinweis auf „Baumgraphen zur Darstellung syntaktischer Interpretationen“ belegen will, so verschenkt er damit aber hundert Jahre Tradition von Baumgraphen in der Sprachwissenschaft, nämlich für Sprachfamilien. Und auch de Saussure und Bühler haben in ihren Werken Visualisierungen benutzt, die bis heute immer wieder zitiert und reproduziert werden.

³ Beispielsweise hatte die Konferenz „Synchrone Betrachtungen, diachrone Erkenntnisse – Slavistisch-Historische Linguistik heute“, die am 5.-6. Dezember 2014 in Freiburg stattfand, eine Sektion zu „Textzeugen vergangener Zeiten im Lichte der Digitalen Geisteswissenschaften“, der Slavistentag im Oktober 2015 in Gießen hatte ein Panel zu diesem Thema, eine weitere Freiburger Tagung fand im gleichen Monat statt.

⁴ Siehe <<http://www.cceh.uni-koeln.de/Dokumente/BroschuereWeb.pdf>>.

menhängen, die eben auch eine räumliche Dimension haben (Wer hat im Mittelalter mit wem von wo nach wo korrespondiert? Wo sind berühmte Leute im Mittelalter geboren worden und wo sind sie gestorben?).

Um sich einen Eindruck von den Methoden der Digital Humanities zu verschaffen, ist es sehr aufschlußreich, einfach in einer Suchmaschine nach Bildern zu dieser Disziplin zu suchen. Dabei erhält man das Ergebnis, das Abb. 2 zeigt (Stand Sept. 2014). Das Suchergebnis zeigt eine Reihe verschiedener Visualisierungstechniken, am häufigsten sticht jedoch eine ins Auge, nämlich die Nutzung sog. 'Wortwolken'. Solche Graphiken scheinen eines der aktuellen Verfahren zur Visualisierung von Sachverhalten in den Digitalen Geisteswissenschaften zu sein, vor allem weil sie einfach zu verstehen sind und im Grunde kaum einer Erläuterung bedürfen.

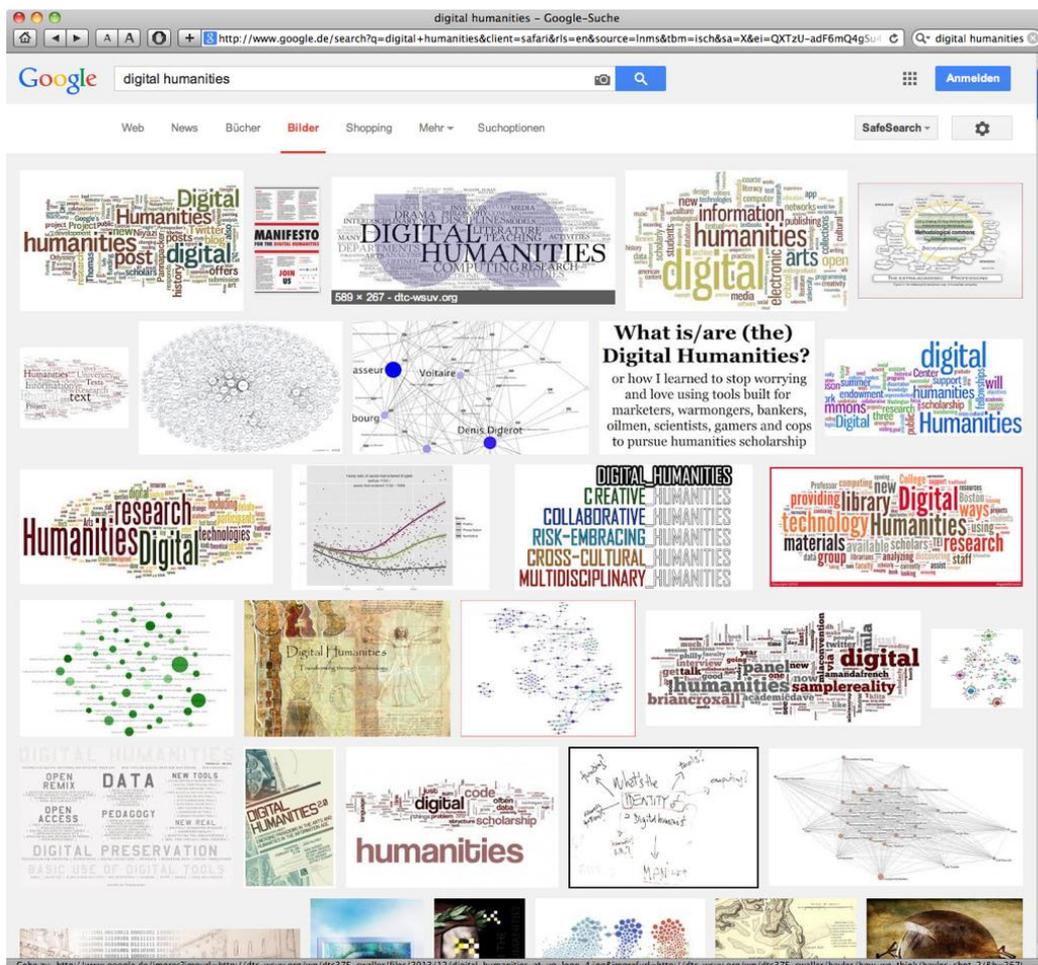


Abb. 2: Suchergebnis für die Visualisierung der „Digital Humanities“ selbst

Ein Beispiel für eine Wortwolke mit linguistischem Gehalt ist das nachstehende (Abb. 3), eine Graphik auf der Homepage eines Lehrstuhls für Angewandte Linguistik (Leipzig). Diese Wortwolke zeigt einige der Stichwörter, die für die Einsatzgebiete der Disziplin stehen, und dabei ganz prominent diejenigen Stichwörter, die nach einer Visualisierung geradezu verlangen: die ‘Korpuslinguistik’ als Beispiel eines ‘Data-driven Turn’ in der Linguistik.



Abb. 3: Wortwolke zur Angewandten Linguistik (Uni Leipzig)

4. Wortwolken erzeugen – ein Beispiel

Wortwolken sind eine mehr oder weniger kompakte, meist horizontale und vertikale Clusterung von Wörtern, wobei die Größe der einzelnen Wörter (genauer: Wortformen) von ihrer Häufigkeit in einem empirischen Ausgangsmaterial bestimmt wird. Produzieren kann man Wortwolken auf verschiedenen Webseiten, wobei wir für den vorliegenden Beitrag die Seite www.wordle.net benutzt haben. Hier kann man in eine entsprechende Box auch lange Texte hineinkopieren – als Testfall diente hier ein achtseitiger Aufsatz des Verfassers (Kempgen 2014). Der Screenshot in Abb. 4 zeigt die Maske mit dem hineinkopierten Text (direkt aus einer formatierten Word-Datei), Abb. 5 das in seiner Aussage überzeugende Ergebnis.

Die hier gezeigte Webapplikation läßt den Benutzer einige Einstellungen vornehmen, wobei wir hier die Voreinstellungen übernommen haben: „Remove numbers“, „Leave Words as Spelled“ und „Remove common English Words“. Das Programm erkennt die Sprache des Dokumentes automatisch, wie die letztgenannte Einstellung zeigt. Zugleich ist diese Web-App auch ein einfaches lin-

guistisches Tool, denn man kann mit ihm die Frequenzliste aller Wortformen des Textes ausgeben – diese muß das Programm ja im Hintergrund berechnen, auf die Liste dann die eingestellten Regeln anwenden sowie anschließend die graphische Ausgabe vornehmen. Für diese gibt es ebenfalls Einstellungen, die vom Nutzer variiert werden können (Schriftarten, Farben, Orientierung, Kompaktheit etc.). Auf diese Details wollen wir hier allerdings nicht eingehen.

Wordle™ Home Create Gallery Credits News Forum FAQ Advanced

Paste in a bunch of text:

A Short Note on the Glagolitic Ornament in Pamvo Berynda's Triod Cvetnaya (Kiev 1631)

Sebastian Kempgen

Although the Glagolitic script is usually associated with Bulgaria and Croatia, graffiti, inscriptions and single letters in otherwise Cyrillic texts or Cyrillic writing areas have been found on East Slavic territory, too – mostly prominently in Kiev and in Novgorod, its first capitals.

In this short note, we would like to draw the readers' attention to two lines of Glagolitic text, which are to be found in the Epilogue to Pamvo Berynda's Triodion or Triod Cvetnaya, printed in Kiev in

Go

OR

Enter the URL of any blog, blog feed, or any other web page that has an Atom or RSS feed.

Submit

Abb. 4-5: Webapplikation zum Erzeugen von Wortwolken (Eingabemaske und Output)

eigenen Forschungsinteressen gleichmäßig auf alle Slavinen verteilt haben. Die Abbildung spiegelt also gut wieder, daß das Russische und seine Analyse den Kernbereich einer Fachvertretung der Gesamtslavistik darstellt.

2. Nach dem Russischen sind die am häufigsten auftauchenden Sprachnamen das Slavische insgesamt sowie das Polnische als Einzelsprache. Dies zeugt einerseits davon, daß der slavische Sprachraum insgesamt vergleichend in den Blick genommen werden soll, ebenso aber auch, daß in der deutschen Slavistik die Westslavistik stärker vertreten ist als die Südslavistik. Außerdem spielt der kontrastive Vergleich mit dem Deutschen die erwartete Rolle. Einzelne weitere, gerade südslavische, Sprachen kommen natürlich vor (Moliseslavisch, Serbokroatisch), aber im Gesamtkorpus nicht sehr prominent.

3. Die Auswertung zeigt mit den Stichwörtern ‘Beispiele’, ‘Problem’, ‘Versuch’ und ‘Überlegungen’ den Arbeitscharakter des Konstanzer Kreises, der es den Vortragenden erlauben soll, Werkstattberichte vorzustellen, Unfertiges zur Diskussion zu stellen usw. Inwieweit die Teilnehmer des Kreises die ersten Überlegungen zu einem Thema später tatsächlich in anderen Publikationen zu Endergebnissen weitergeführt haben, wäre eine andere Frage.

4. Inhaltlich stehen Syntax und Semantik sowie der Aspekt (Stichwörter ‘Aspekt’, ‘Aspekts’ und ‘Verbalaspekt’) erwartungsgemäß ein wenig im Vordergrund, vom linguistischen Ansatz her fallen die Stichwörter ‘Funktionen’ und ‘Kategorien’ auf, während ein denkbare Stichwort wie ‘generativ’ fehlt, da dieses Grammatikmodell unter den Teilnehmern des Kreises nicht stark vertreten war bzw. ist.

7. Konstanz und Variation bei Individuen / Autorschaftserkennung

Die vorliegenden Daten erlauben natürlich auch längsschnittliche Auswertungen zu den Forschungsschwerpunkten einzelner Personen. Bei der Tagung des Kreises wurde diese Frage umgedreht, so daß damit ein wenig getestet werden konnte, ob Wortwolken tatsächlich ein heuristisches Mittel sein können, um Autorenerkennung zu leisten, Urheberschaften erkennen zu können usw. Es wurden vier Auswertungen für jeweils alle Vortragstitel einer Person vorgelegt, und die Anwesenden sollten den Autor erkennen. Dies gelang ohne Schwierigkeiten, unter anderem deswegen, weil der Kreis der Infragekommenden ja auf den Konstanzer Kreis selbst beschränkt war. Auf jeden Fall ließen die vier separaten Auswertungen aber erkennen, daß manche Slavisten und Slavistinnen eine größere Konstanz in der Thematik ihrer Publikationen aufweisen als andere, wohingegen andere die Vielfalt bevorzugen.

Eine solche Auswertung soll hier beispielhaft gezeigt werden – vgl. Abb. 9. Sie demonstriert zugleich das bekannte Phänomen, daß man Individuen nicht an dem erkennt, was vielen gemeinsam ist (Stichwörter ‘russisch’, ‘Funktion’),

zu den Orten, an denen in Deutschland Slavistik-Standorte zu verzeichnen sind, so ergibt sich die nachstehende Karte (Abb. 10).

Anhand dieser Karte lassen sich mehrere Dinge resümierend festhalten:

1. Der Konstanzer Kreis umfaßt leicht die Hälfte aller wichtigen slavistischen Standorte im deutschsprachigen Raum (alle nicht grauen Kästchen); wichtige Ausnahmen sind vielleicht die Standorte Leipzig und die HU Berlin, die beide große Slavistiken haben. Dabei muß man beachten, daß der Kreis seine Mitgliederzahl bewußt begrenzt, man also nicht einfach beitreten kann.

2. Nur zwei Orte, an denen der Konstanzer Kreis Mitglieder hatte/hat, sind je geschlossen worden oder von Schließung bedroht gewesen (Frankfurt, Bayreuth; roter Rand um leeren Kasten).

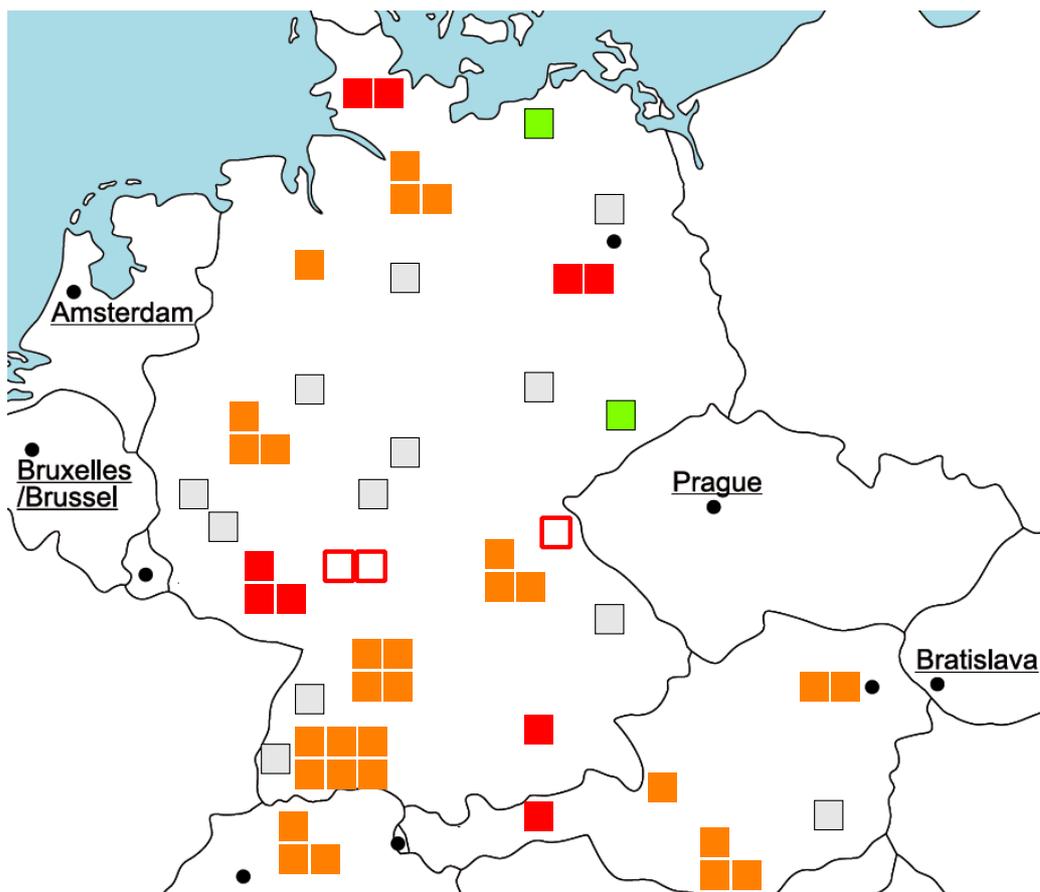


Abb. 10: Tagungsorte des Konstanzer Kreises im slavistischen Kontext

Es soll hier zwar kein ursächlicher Zusammenhang behauptet werden, aber die Daten zeigen vielleicht doch, daß Netzwerkbildung immer seine 'Knoten' stützt

und insofern wirkt, bzw. es zeigt sich, daß die Mitglieder des Konstanzer Kreises auch sonst für ihren jeweiligen Standort sichernd gewirkt haben. ‘Netzwerkbildung’ als Rezept für kleine Fächer ist also ein probates Mittel, dessen man sich bewußt bedienen sollte.

3. Das Slavistische Arbeitstreffen wurde in Westdeutschland ins Leben gerufen, und ähnlich wie im Deutschen Slavistenverband insgesamt war es eine Herausforderung der Gruppe, die Arbeitskontakte und Netzwerkbildung auf die Standorte der ehemaligen DDR auszudehnen. Das gelang früh in der Person von Rudolf Růžicka und setzt sich aktuell wenigstens in Bezug auf zwei ostdeutsche Standorte der Slavistik fort (Dresden, Greifswald, grüne Kästchen).

4. Biographien von Wissenschaftlern (und Wissenschaftlerinnen) bringen Ortsveränderungen mit sich, und deshalb hat die Ausdehnung der Tagungsorte natürlich auch mit Berufungen einzelner Mitglieder zu tun. Bei einer Handvoll Standorte ist erfreulicherweise inzwischen eine Mitgliedschaft schon ‘in 2. Generation’ zu verzeichnen (Beispiele: Konstanz, Bochum, Tübingen, Hamburg, Klagenfurt), wohingegen es umgekehrt aber auch Standorte gibt, bei denen das 21. Jahrhundert keine Tagung mehr verzeichnet und eine solche auch nicht unmittelbar zu erwarten ist (Kiel, Mainz, München, Innsbruck: rote Kästchen), wobei dies mehrheitlich Standorte sind, von denen im Konstanzer Kreis niemand Mitglied ist. Das Bild ist also einerseits erfreulich, weil es die Attraktivität der Zugehörigkeit zu diesem Netzwerk beweist, andererseits zeigt es, daß Kooperationen weder verordnet werden können noch sollen, noch daß die schulenmäßig oder sonstwie bedingte ‘Passung’ zu Netzwerken automatisch gegeben ist.

5. Die Karte zeigt auch, daß der Konstanzer Kreis bewußt ein deutschsprachiges Arbeitstreffen geblieben ist, auch wenn zu einzelnen Sitzungen internationale Gäste eingeladen wurden. In der Slavistischen Linguistik haben sich für den Nachwuchs andere Netzwerke etabliert, in denen diese Frage von vornherein anders gehandhabt wurde; Alternativen sind also gegeben.

Jeder Arbeitskreis, der wissenschaftlich über die Diskussion unter seinen Mitgliedern hinaus wirken möchte, wird natürlich seine Rezeption im Auge haben. Mit diesem Aspekt hatte beim Konstanzer Kreis vor etlichen Jahren die Frage der Publikationsform und -disziplin zu tun, die im Hinblick auf die Verbreitung nach Sammelbänden nunmehr zugunsten einer Zeitschrift entschieden wurde. Inzwischen bieten hier Portale wie *Academia.edu* oder *ResearchGate.net* neue Möglichkeiten, Pre- oder Postprints seiner Veröffentlichungen auf der Basis der Open Access-Gedanken einem potentiell viel größeren Leserkreis zur Verfügung zu stellen, und auch slavistische Zeitschriften tolerieren oder fördern solche Verbreitung ja in immer größerem Maße.

Literatur

- Kempgen, S. 2014. A Short Note on the Glagolitic Ornament in Pamvo Berynda's *Triod' Cvetnaja* (Kiev 1631). Preprint, published electronically, University of Bamberg, 8 pp.
- Rathmayr, R. 2000. „Mir scheint, daß es im Augenblick mehr Fragen als Antworten gibt.“ Eine quantitative Analyse des wissenschaftlichen Schreibens anhand von 279 Schlußsätzen und 25 Jahren Konstanzer Kreis. In: Breu W. (ed.) *Slavistische Linguistik 1999. Referate des XXV. Konstanzer Slavistischen Arbeitstreffens Konstanz, 7.-10.9.1999* (= Slavistische Beiträge 396). München, 187–209.
- Schleicher, A. 1861. *Compendium der vergleichenden Grammatik der indogermanischen Sprachen*. I. Weimar.

