

On the usefulness of Google Books & co. for histories of concepts (not only in the Slavic Studies): *glasnost'*, standard language, and digraphia/bigraphism

The article shows how new insights for the history of concepts can be gained using Google Books and other large collections of full texts. Various problems that are encountered are discussed. By demonstrating the method, new light is shed on the history of the concepts of *glasnost'*, standard language and digraphia/bigraphism.

Translated from "Über den Nutzen von Google Books & Co. für (nicht nur slavistische) Begriffsgeschichten" in Podtergera, Irina (ed.), *Schnittpunkt Slavistik: Ost und West im wissenschaftlichen Dialog; Festgabe für Helmut Keipert zum 70. Geburtstag*, vol. 1, 393–407. Göttingen: V&R unipress – Bonn University Press 2012. DOI: 10.18716/bun/gg1b.

DOI: 10.18716/omp.64.5

One of Helmut Keipert's many contributions to the history of Slavic words and terms,¹ the essay on *glasnost'* published shortly before his retirement, begins as follows:

„Die folgenden Ausführungen [...] sollen vor allem daran erinnern, daß man bei begriffsgeschichtlichen Studien in der Russistik von der Lexikographie nicht diejenige Unterstützung erhält, die dafür erforderlich wäre“ (Keipert 2006a: 1).

“The following remarks [...] are primarily intended to remind readers that for studies on the history of concepts, Russian lexicography does not provide the support that would be necessary for them.”

However, help may now be available from a completely different source—namely, databases of digitized full texts. In this short paper, I would like to explore how this works, how much support they can provide, and what they cannot accomplish, using a few examples.

Among the full-text databases currently available, the Google Books project run by the American company Google Incorporated is by far the most ambitious and advanced. Among other things, the plan is to scan the entire library holdings

¹ The systematic list of writings published in the festschrift for his 65th birthday (Bunčić 2006) contains as many as 15 papers devoted to the history of a single word or term (positions 27, 58, 115–127). At least two more recent papers should now be added to this list, namely another on *cerkovnoslavjanskij* (Keipert 2006b) and one on *narodnost'* (Keipert 2008).

of the University of Michigan. In addition to many American libraries, several European libraries (and one in Japan) have now also decided to collaborate with Google, which means that the historical collections of the Austrian National Library, which are no longer protected by copyright and are so important for Slavic Studies, as well as the Bavarian State Library, which held the Special Subject Collection for Slavic Studies in Germany (funded by the German Research Foundation) until 1997, are now also to be digitized. Unfortunately, □ 394 no library in a Slavic country has joined the project yet (as of April 2011; <https://web.archive.org/web/20110806014418/http://www.google.de/googlebooks/partners.html>). Nevertheless, a considerable number of Slavic books—from the collections of the participating American and European libraries—are already available via the book search.

In the entire collection of books scanned to date, you can search for words, phrases, and exact word forms (in quotation marks) in the usual way using the Google internet search engine (at <https://books.google.com/>). Unlike an Internet search, however, the advanced search form allows you to limit your search to a specific period and display the results in chronological order (unfortunately only in reverse chronological order, as the technicians working on this project apparently did not think of the possibility of searching for the oldest reference instead of the most recent hit). If the copyright for the book found has expired, the page on which the hit is located is displayed as a facsimile, with the words searched for usually highlighted in color, and often you can even download the entire book as a PDF file if you wish. In the case of newer books that are still under copyright, sometimes only individual pages are displayed (and the hit may or may not be on one of the pages displayed), sometimes you only get a small “snippet” with the page number, and sometimes even just the general information that a particular book contains the word. However, even such sparse information can be helpful in finding the desired reference in a library in the traditional way.

Since only keywords can be entered in the book search engine, this tool is only suitable for those types of the history of concepts that take a “middle path between word history, factual history, and problem history” (“vermittelnder Weg zwischen Wortgeschichte, Sachgeschichte und Problemgeschichte”, Keipert 2006a: 4) and thus have a semasiological element. An onomasiological search for expressions with a specific meaning is, of course, not possible. However, in the case of histories of concepts that trace the development of different expressions for a term, full-text databases can also help to illuminate the history of each of these expressions if the expressions used for the term are known from other sources.

As our first search term to illustrate the method, let us take the keyword *glasnost*, which became known internationally through Gorbachev and which Keipert (2006a) uses as an example to criticize the inadequate representation of the history of the term in Russian dictionaries. Even in the 19th century, the □ 395

keyword, which some consider a neologism of the 1980s, played a political role when it came to the *publicity* of court proceedings in particular (*glasnost' sudoproizvodstva*). The word itself is, of course, somewhat older; the first known evidence of its use can be found in Trediakovskij (1752). This is also the oldest hit that Google Books brings up (<https://books.google.de/books?id=s0oEAAAAYAAJ&pg=PA9>, digitized on 3 Jul 2007²). There, however, *glasnost'* is used only in the sense of “pure pronunciation, clear articulation” (“reine Lautung, klare Aussprache”, Keipert 2006a: 7, 12). The central problem, which even Keipert (2006a: 13f.) was unable to solve with the means available at the time, is precisely that “we cannot yet trace the path of metaphorization of *glasnyj* from ‘loud’ to ‘openly apparent, generally known’” (“wir den Weg der Metaphorisierung von ‘laut’ zu ‘offen zutage liegend, allgemein bekannt’ bei *glasnyj* bisher nicht verfolgen können”).

Here, Google Books really seems to be able to help. Three hits from a search for *glasnyj* are particularly interesting: the first is from a dictionary that Keipert (2006a: 13) also quotes—in a later edition—the *Polnyj německo-rossijskoj leksikon, iz bol'sago gramatikal'no-kritičeskago Slovarja gospodina Adelunga sostavlennyj* (vol. 2, Sanktpeterburg 1798). However, Keipert searched for *glasnyj* as a counterpart to *öffentlich* ‘public’ and did not find it in that article. The digital full-text search has the great advantage that one does not have to know in advance where one might find something. One would not have readily looked up *ruchtbar* ‘notorious’, but this very adjective is translated as “извѣстный, вѣдомый, гласный, явный” and illustrated, among other things, by the example “Eine ruchtbare That, гласное дѣло” (“a notorious act”). The corresponding noun can also be found here: “die Ruchtbarkeit, (plural: неуп:) извѣстность, явность, гласность” (...?id=Dc0GAA AAQAAJ&pg=PA320, 2 Mar 2007).

The precise usage of *glasnyj* at the end of the 18th century is illuminated by a letter from Catherine the Great to her emissary at the Sublime Porte, Jakov Ivanovič Bulgakov, dated 30 December 1782, which is reprinted in Dubrovin (1889: 974–976; and this collection of sources from 1781 and 1782 is sensibly filed under 1782 and not 1889 on Google Books). At the end of lengthy instructions on diplomatic maneuvers, the tsarina writes:

«Покуда Порта о семъ не станеть вы- зываться, вы сохраните все сіе въ непроницаемой тайнѣ, да и нѣтъ нужды сообщать сіе предварительно **Д 396** ин- тернунцію вѣнскаго двора; но когда

“Until the Porte raises this issue, keep all this in strict *secrecy*, and there is no need to inform the internuncio of the Viennese court in advance; but when it already *becomes public knowledge*, then you may

² In the following, the addresses of the documents on Google Books are shortened by the constant “<https://books.google.com/books>”, so that only the identification number and the reference to the page with the hit are given, supplemented by the date of digitization. The retrieval date is generally omitted—all information relates to April 2011 [and has been updated to March 2026 for the English translation—D. B. 2026].

уже оно *учинится гласнымъ*, тогда вы можете сказать интернунцію о вопросахъ, отъ Порты вамъ сдѣланныхъ, и о вашемъ отвѣтѣ [...]» (ibid. 975f.; ...?id=GSYbAAAAYAAJ&pg=PA976, 23 May 2008; my emphasis).

inform the internuncio of the questions put to you by the Porte and of your reply [...]”

In a similar context, Levašov (1790: 155f.) uses the word in a report about his imprisonment in the Ottoman Empire:

«[...] Апреля 1го числа явился у насъ [...] вѣстникъ съ объявленіемъ, что мы будемъ отпущены въ отечество [...], и чтобы мы сіе содержали до времени въ тайнѣ, дабы народъ несталъ на Правительство роптать [...]; но осторожность сія не была сохранена въ Царь-градѣ съ надлежащею точностію, гдѣ немѣленно разпространился слухъ, что насъ выпустятъ на волю [...], и мы увѣдомясь чрезъ пріятелией своихъ, что освобожденіе наше вездѣ почти *содѣлалось* уже *гласнымъ*, опасались, чтобъ сіе самое неоставило насъ и надолго въ Демотикѣ [...].» (...?id=07o9AAAAYAAJ&pg=PA155, 8 Jan 2009; my emphasis).

“[...] On April 1, a [...] messenger arrived with the announcement that we would be released to return to our homeland [...], and that we should keep this information *secret* for the time being so that the people would not begin to complain to the government [...]; but this caution was not maintained with due precision in Constantinople, where rumors immediately spread that we would be released [...], and when we learned from our friends that our release had already *become public knowledge* almost everywhere, we feared that this very fact would keep us in Demotika for a long time [...]”

In both cases, *glasnyj* is used to refer to a secret that has *become public knowledge*—something that should have been *kept secret* is now obvious. This could be the missing link between the original and the transferred meaning: when a secret is no longer a secret, it has usually been *spoken about*, *gossiped* or even *announced*; in any case, the silence has been broken by a voice (*glas*). This development in meaning would be parallel to the German *ruchbar* (*werden*), because this goes back to Middle Low German *ruchte* ‘call, cry, reputation’ (and, like *Gerücht* ‘rumor’, *berüchtigt* ‘infamous’ and also *anrühlich* ‘objectionable’, not to *riechen* ‘to smell’, cf. Pfeifer 2003: s.v.). From there, it is only a small step to the meaning ‘public’, because when *glasnost*’ became a buzzword in the 19th century, it was, as Keipert (2006a: 9) rightly notes, a “future concept” (“Zukunftsbegriff”), i.e. the court proceedings were still secret at that time. What was being negotiated in secret was therefore yet to *become public knowledge*. Only when *glasnost*’ is applied to an already existing practice does its meaning approach that of *publičnost*’, *javnost*’ ‘publicity’. Significantly, however, Gorbačëv’s use of this keyword was also about the *introduction* of freedom of information and expression, so here, too, something that had previously been secret was now to be spoken about aloud.

The fact that not only in Russian or, as Keipert (1998: 122) has stated elsewhere, “in most Slavic languages the lexicographical prerequisites for more extensive etymological studies are lacking” (“für weiter ausgreifende wortgeschichtliche Studien in den meisten slavischen Sprachen die lexikographischen Voraussetzungen fehlen”) but that German and English dictionaries, for example, [□397](#) also leave room for improvement, particularly with regard to technical terms, can be seen in the history of the term *standard language*. Gröschel (2009: 92–95) has traced this history on the basis of specialized and general dictionaries: While Joseph (1987: 5) finds the first attestation of *standard language* in English in 1858, dictionary research shows that *standardni jezik* has been documented in Serbo-Croatian since 1964, *literaturnyj (standartnyj) jazyk* in Russian since 1966, and *Standardsprache* in German since 1968. Gröschel (2009: 93) supplements the evidence from Russian dictionaries with a reference in the title of a much earlier paper by Polivanov (1927). Gröschel (2009: 93) has “more than just slight doubts” (“mehr als nur gelinde Zweifel”) about this chronology, especially since he cannot imagine that “such a large time gap should yawn” (“eine derart große zeitliche Lücke klaffen sollte”) in German between *Standard*, which was already documented in 1804, and *Standardsprache*, which was first documented in 1968 (ibid. 94).

As Google Books shows, these doubts are entirely justified. For example, *Standardsprache* can already be found in several journal articles from the end of the 19th century, first in two reviews by Siebs (1897a: 220, ...?id=18cqAAAAMAAJ&pg=PA219, 12 Feb 2007; 1897b: 555, ...?id=A-levztN2O8C&dq=Standardsprache&pg=PA555, 4 Jun 2007; Siebs’ orthoepic dictionary, on the other hand, still exclusively writes *Hochsprache* for ‘standard language’ in its 16th edition of 1957). Jespersen (1904: 39) also refers to “Einheitssprache, Gemeinsprache, Standard-sprache oder wie man sie sonst nennen will” (“uniform language, common language, standard language, or whatever else you want to call it”). Up to 1967, Google Books yields a total of 146 books containing the term *Standardsprache*, which clearly shows that the term was used more than sporadically before it first appeared in a dictionary. In Serbo-Croatian, references in journals can be found from 1950 onwards (first in Tabak 1950: 320; ...?id=FwVBAQAIAAJ&q=standardnim). For Russian, Polivanov (1927) has not yet been digitized, but bibliographic references to this essay can be found, as well as a number of attestations of *standartnyj jazyk* between 1927 and 1966.

The first recorded use of the English term *standard language* in its modern sense³ can also be traced back to 1858, at least 81 years earlier than previously

³ In an obviously different sense, probably as a nonce word, *standard language* appears in Fell (1729: 302), in the report on the events following the capture of the city of Hippo by the Vandals in 431, one year after the Church Father Augustine had died there: “[...]but the Barbarians shew’d Respect to the Saint’s Body and Writings, which made it evident that the Almighty refrain’d them from infulting the Remains of His Servant, and refered

thought. In Google Books, [▢398](#) a review of Clarke (1777) can be found in *The Critical Review: or, Annals of Literature* 45 from 1778 (pp. 248–252; ...?id=7cI-PAAAAQAAJ&pg=PA248, 13 Mar 2008) in which the book is quoted. Although the book itself has not been digitized by Google Books, it is available online in all German university libraries through a national license. It deals with the education of young men, and in this context the author explains:

“If such attention is also given to the dialect and pronunciation, as to form their language to the *national standard*, they may one time or other be benefited thereby; especially if they have occasion to be in places distant from that of their nativity. I have seen men who knew not *p* from *q*, by being habituated in youth to hear and speak the *standard language*, have more address and sentiment, than others who had been taught reading, writing, and the use of numbers [...]” (Clarke 1777: 18; my emphasis)

A document written five years later makes it clear that the word is also used in philology itself. In the dispute over the authenticity of the 15th-century poems by Thomas Rowley, which were forged by Thomas Chatterton in the 18th century (there was also an “invented past” (“*erfundene Vergangenheit*”, cf. Keipert 2001) in Western Europe), Tyrwhitt (1782: 3) opposes two arguments put forward by those who believe in the poems’ authenticity:

“1. That the Poems are written in a provincial dialect, and therefore are not reducible to the rules of the *standard-language*. 2. That there was no *standard-language* in the XV century, by which they can be tried.” (...?id=JkZWAAAAYAAJ&pg=PA3, 3 Mar 2011; my emphasis)

Here, too, Google Books is more helpful than traditional lexicography.

Highly specialized scientific terms, such as those describing the phenomenon of **biscriptality**—i.e. the fact that several scripts are used for one and the same language (which is the subject of my postdoctoral research project)—are hardly ever included in dictionaries. Of particular interest in terms of the history of the concept is the term *digraphia*, which is based on diglossia according to Ferguson (1959) and which was coined independently several times. It was popularized by DeFrancis (1984), who notes in a footnote that after completing his manuscript, he learned that this term had already been used in a similar sense by Dale (1980). However, Grivelet (2001: 1f.) cites Zima (1974) as “[t]he first discussion of the notion of digraphia”, and Unseth (2008: 3) adds Jaquith (1976) as another “inventor” of this term. A Google Books search [▢399](#) for “digraphie” finds Gebhardt

his Works as the Standard Language of the Church on the important and difficult Articles of Grace and Predestination” (...?id=CRgQ6lqsuOwC&pg=PA302). There, *language* is still to be understood in a sense that is no longer common, which the *Oxford English Dictionary* (s.v. *language* 5a) describes as “That which is said; talk, report, rumour”, here perhaps more freely translated as *doctrine* or *teaching*.

(1974: 175; ...?id=CpIrAAAAMAAJ&q=digraphie), among others. Unfortunately, this book can only be viewed in “snippet view”, so one must first borrow it from a library to see that Gebhardt here refers to Lafont (1971). Lafont’s article is freely accessible on the journal portal *Persée* (doi:10.3406/lfr.1971.5576⁴), and so arguably the very first evidence of the linguistic term *digraphie* for ‘biscriptality’ has been found, which Lafont mentions only in passing and which was certainly not taken up by the other “inventors”. This passage discusses the fact that Occitan not only competes with Standard French, but also has two competing orthographies:

« La situation de diglossie occitane n’est donc pas semblable absolument à celles qu’on peut trouver en d’autres lieux de contacts linguistiques : elle se complète par une situation de digraphie » (Lafont 1971: 95).

“The situation of Occitan diglossia is therefore not at all similar to those found in other areas of language contact: it is complicated by a situation of digraphia.”

As an aside, let me remark that the *Bibliographie linguistique*, whose online edition, after a free trial and promotional period, is now unaffordable for many libraries, can be viewed in a limited preview on Google Books, so that a keyword search for *digrafia* there leads to references to the second and third parts of Consani’s (1988, 1989, 1990) trilogy of articles under nos. 7030 and 7031 in the *Bibliographie linguistique* for the year 1990 (...?id=os9TMERY9OsC&pg=PA357). This allows us to trace a fifth and (so far) last independent “inventor” of this expression.

Similarly interesting and complex is the prehistory of sociolinguistic concepts of biscriptal writing, namely the emergence of expressions for ‘biscriptal’, e.g. for documents that contain the same text in two scripts (but in the same language). This has not yet been researched at all. Google Books finds the oldest relevant mention of *digraphic* in Pierides (1876; ...?id=IBsXAQAIAAJ&q=digraphic), but only small excerpts are displayed. If you then obtain the article from the library, you will find that Pierides (1876: 38) actually acknowledges having invented the term himself:

“In the summer of 1873 I became possessed of an inscription in Greek and Cypriote, then discovered in Larnaca, the ancient Citium. [...] As the language is the □ 400 same in both parts, and only the writing differs, I prefer calling this inscription *digraphic*, instead of *bilingual*, until a better definition is proposed.”

Like so many temporary solutions, this one also proved to be surprisingly durable, as the term is still in use today, especially in Greek philology—Consani (1988, 1990) undoubtedly continues this usage with his coinage of *digrafia*. At almost the

⁴ Unlike the *Uniform Resource Locator (URL)*, the *Digital Object Identifier (DOI)* is permanent. If the browser does not automatically resolve such addresses, “doi:” has to be replaced with “https://doi.org/”.

same time, von Sallet (1875: 132) independently introduced the German term *zweischriftig* in a completely different discipline, numismatics, for coins that contain inscriptions in two different scripts—here, too, the two scripts are the Cypriot syllabary and the Greek alphabet:

„[...] einige dieser Münzen, welche als zweischriftig – sit venia verbo – besonders interessant sind, geben neben der cyprischen auch die griechische Legende [...].“

“[...] some of these coins, which, being *zweischriftig*—if you permit the expression—, are particularly interesting, also bear the Greek inscription in addition to the Cypriot one [...].”

This hit can also only be viewed as a “snippet” on Google Books (...?id=CU1VIO6izyEC&q=zweischriftig, 17 May 2005) and had to be obtained in paper form for the complete text. Another synonym for *digraphic* and *zweischriftig* is *bigraphic*, which is documented much earlier than the other two, namely by Pierquin de Gembloux (1840). However, the content of this book is more than dubious, as it describes, as a preliminary study for a *History of the Fatherland before the Roman Conquest* (*Histoire de la Patrie avant la conquête romaine*, *ibid.* X), the migrations of the Celts across half the world, including America. For example, on a large rock in the Mississippi there allegedly is “a bigraphic Celtic inscription, i.e. half hieroglyphic, half alphabetic” (« une inscription celtique bigraphique, c’est-à-dire moitié hiéroglyphique, moitié alphabétique », *ibid.* 248; ...?id=jh1GAAAACAAJ&pg=PA248, 25 Jan 2007). Here again, the advantage of the method becomes apparent, because no one would have thought to look for the first evidence of this linguistic term in such an abstruse work. Of course, it is also unlikely that this work influenced the later use of this term in linguistics, so that we are probably dealing here with a nonce coinage. Later users of the same term, including Siméon (1889: IX), who comments on Mexican pictographic manuscripts with Latin glosses in Nahuatl, saying that they “could more exactly be called *bigraphic*” (« pourraient être plus exactement appelés *bigraphiques* »), are certainly independent of Pierquin. For the latter book, however, Google Books only provides a reference that it contains the word sought (...?id=3V8SAAAAYAAJ&pg=PR9, 5 Dec 2007), after which one must borrow it in paper form and search for the relevant passage. However, the address referred to in the hit list contains a code for the page number, even if it is not displayed on the page (“PR9” in this case is p. IX, with ‘R’ for Roman numerals, as opposed to ‘A’ for Arabic numerals).

▮401 The extent to which the “forgetting of knowledge” is not only commonplace “in Slavic philology” (“Vergessen von Wissensbeständen [...] in der Slavischen Philologie”, Keipert 2006c) but also in other fields, is demonstrated, apart from the sixfold ‘invention’ of *digraphia*, also by the fact that Blake (1995: 463) still speaks of a “‘bigraphic’ scribe (to invent a suitable term)” (“escriba ‘bígrafo’ (para inventar un término adecuado)”, ...?id=sq7jAAAAMAAJ&q=bígrafo, “snip-

pet view”), even though, as seen above, this term has not needed to be invented since 1889 (or 1840).

One of the earliest examples of a noun describing ‘biscriptality’ (of texts), incidentally, comes from Russia, specifically from Ol’denburg (1899: 208), who writes about an Indian manuscript:

«[Ч]резвычайно любопытную особенность этого отрывка составляет то, что въ немъ мы имѣемъ образчикъ *биграфизма*, а именно на листѣ 27b. мы встрѣчаемъ *одновременно* и письмо характера *индійскаго* *gupta* и *кашгарскаго* [...]» (original emphasis)

“An extremely curious feature of this passage is the fact that it provides us with an example of *bigraphism*; namely, on folio 27b, we encounter *simultaneously* writing in both the *Indian* *gupta* script and the *Kashgar* script [...]”

However, this text has not yet been included in Google Books. Nevertheless, when searching for *Bigraphismus*, one comes across a “snippet” from Barthold (1899: 140; ...?id=byfWAAAAAAAJ&q=bigraphismus&pg=PA140, 29 May 2009), which, as one can discover by examining a paper copy of this journal, reports on Ol’denburg’s essay in German. In this way, despite Google Books’s focus on Western libraries, so far neglected Slavic works sometimes leave their mark.⁵

These brief excerpts from the history of the concept of biscriptality may suffice to demonstrate the possibilities but also some of the **obstacles** of working with Google Books. Here, the impossibility of an onomasiological search proves to be a serious problem, because the multitude of terms that have been used for the phenomenon of biscriptality (in addition to *digraphia*, *bigraphism*, and *Zweischriftigkeit*, there are also *bialphabetism*, *biscriptalism*, *orthographic diglossia*, and others) must be entered individually into the search engine—and in all languages in which relevant evidence is hoped to be found. In doing so, one can never be entirely sure that one has not overlooked an important discussion that could actually be found with the available means, because another term was used there that one happens not to know.

▣402 Almost as large a problem as the synonyms used for a concept to be investigated is posed by polysemy and homonymy. For example, the relevant hits for a search for *digraphie* have to be fished out of a sea of irrelevant hits, since *digraphie a*) is also used as an alternative form of *digraph* (referring e.g. to Polish *cz*, *ch*, *rz*, etc.), b) refers to a method in radiology used to display the lungs in both

⁵ A much more detailed treatment of the history of this term will be published as part of my habilitation thesis. [It was published as chapter 2 of Bunčić (2016: 27–50), which elaborates on the history of the notion itself but does not say much about the methods used to investigate this history. — D. B. 2026.]

the inhaled and exhaled state on the same X-ray image, and c) is the usual word for double-entry bookkeeping in French.

Moreover, typing and reading errors can lead to incorrect hits. When searching for *bigraphisch* ‘bigraphic’, for example, the vast majority of hits are misspellings or incorrect readings by the text recognition program for *biographisch* ‘biographical’ (because if the *o* could not be recognized correctly, something like *bi•graphisch* is returned as a hit for *biographisch*). Poor image quality can lead to absurd errors—for example, Google Books reads the following passage in Schneidewein (1740: 466)

Ut quod in curia mercatorum non obtineat

as “Ut quod in curia гласности non brincar” (...?id=mcnXB3uy4CcC&pg=PA466, 13 Jan 2010; with my human eyes, I read “Ut quod in curia mercatorum non obtineat”). Another problem is word separation at the end of a line, which means that a search for *glasnyj* also returns a number of hits where *glasnyj* appears at the beginning of a line, but the previous line ends, for example, with *so*.⁶

When searching for the use of a word in the past that is very common in the present, it is often annoying that the metadata, including the publication year, always refer only to entire volumes. Thus, one is surprised to find a word such as *Standardsprache* ‘standard language’ in a text from 1484 (...?id=AMwJAQAIAAJ&q=standardsprache, 11 Feb 2009)—but of course this is not the oldest attested use of the term in German but a hit in the accompanying text of a scholarly edition (in this case of Stephan von Landskron’s *Hymelstrasz* of 1484, published by G. J. Jaspers in 1979 “mit einer Einleitung und vergleichenden Betrachtungen zum Sprachgebrauch in den Frühdrucken”, i.e. “with an introduction and comparative observations on language use in early prints”). In addition, there are a surprising number of typing errors in the metadata (or are they also obtained through automatic text recognition?). One example is what at first glance appears to be the earliest evidence of *standard language* in English in a text dated to 1708 (where, incidentally, the term is used in the sense of standardized English-language biological terminology). However, this turns out to be a letter to the editor from the *Monthly Magazine* of February 1798 (...?id=arXaiokRtygC&pg=PA110, 21 Sep 2005).

▣ 403 Google Books is controversial primarily because of its handling of **copyright**, and for good reasons. The opt-out procedure sought by Google, whereby the author (or copyright holder) of a printed work must be aware that Google is scanning

⁶ As of 2026, the first two of these problems have been solved, probably by improved text recognition; the passage from Schneidewein is now contained as “Utquod in curia mercatorum non obtineat” in Google Books. False positives due to hyphenation are also much less common now. (D. B. 2026.)

this work in order to object to it, is certainly not ideal from the perspective of commercial interests. Admittedly, though, the most comprehensive possible collection of printed literature, as intended by Google and desirable for scientific purposes, will never be achieved if all authors who have not been dead for more than 70 years must give their express consent (*opt-in* procedure). It seems to me that a slightly expanded “snippet view” (which, especially in the case of journals and anthologies, would have to contain the bibliographic information of the individual article) would be a sensible compromise for all works that are not in the public domain or have been released by their copyright holders. This would not harm the authors but would provide researchers with enough information to decide whether a text contains relevant information and, if necessary, should be obtained by other means. However, this issue seems to be of little relevance to the cases discussed here, as the hits all come from works whose copyright has either expired or which have been made available by the publisher on the basis of an agreement with Google (in these cases, the date of digitization is usually missing).

A much bigger problem than the copyright dispute is Google’s monopolistic position. It is highly questionable whether a single private company, which is not subject to any government or intergovernmental control, should be allowed to make a large part of the world’s knowledge universally accessible (even if this company has the unofficial motto “Don’t be evil”). Therefore, **alternative projects** are very welcome. The most comprehensive such institution besides Google Books (which, according to Armstrong, has scanned 15 million books to date) is *HathiTrust* (<https://www.hathitrust.org/>), a very promising network of American research libraries that offers not only a union catalog but also a unified search of all digital full texts available at these libraries—including those digitized as part of the Google Books program. Unfortunately, there is no advanced full-text search function yet, which means that searches in the digital documents cannot be limited in time, filtered, or sorted, making this tool unsuitable for the purposes of conceptual history at present.

▣ 404 *Europeana*, an association of European libraries, museums, and archives, according to its own information, contains “more than 15 million objects” (including images, audio recordings, and videos, <https://europeana.eu/about-us>). Unfortunately, there is no functional full-text search feature yet. However, as with HathiTrust, we can expect improvements in the near future, which will hopefully make these databases genuine alternatives and additions to Google Books.

Gallica, the digital library of the French National Library (<https://gallica.bnf.fr/>), has proven helpful in several searches. It currently has 1.5 million digitized documents, but these are almost exclusively French texts and texts related to France. A Polish equivalent is the *Federacja bibliotek cyfrowych*, which currently has over 600,000 digitized publications (<https://fbc.pionier.net.pl/>), but so far it is only

possible to search the metadata. The same applies to the *Ėlektronnaja biblioteka* of the Russian State Library in Moscow, which has digitized 630,000 dissertations and over 8,300 old prints, among other things (<http://elibrary.rsl.ru/>).

A project with a different focus is JSTOR, which specializes in digitizing older issues of scientific journals (<https://www.jstor.org/>). Although the organization itself is non-profit, its services are so expensive that most libraries only purchase individual sub-collections. Nevertheless, it is possible to perform a full-text search of the entire collection; for articles that are not subscribed to by your own library and that contain the search term, only their bibliographical information is displayed. (Some articles can also be purchased directly online—according to my random samples, at prices ranging from \$9 to \$38.) However, searches conducted in JSTOR in parallel with the Google Books searches described above revealed that most of the relevant information could not be found there. In addition to the deliberate restriction to journals, this is probably also due to the strong focus on English-language sources.

Overall, I hope to have shown that Google Books can be useful for researching the history of concepts (and other full-text databases will certainly be useful in the future as well). Even with this tool, reading the original paper texts will not become unnecessary in the foreseeable future, and this method also requires a certain overview of the use of the term being investigated. In this respect, book searches cannot be compared to using Google's internet search engine as a corpus, which Robb (2003) describes as "quick 'n dirty". Since you often only get text snippets or bibliographic information and □405 then, in order to read the entire text, may have to resort to traditional interlibrary loans, searching via Google Books is not necessarily fast. On the other hand, the results obtained in this way are by no means unreliable (as long as one is aware that there may be other uses of the search term that are not covered by Google Books). Therefore, given the often inadequate representation of the history of terms in dictionaries (not only Slavic ones), the method presented here can be a valuable aid.

References

- Armstrong, Kim. 2011. One million books scanned and returned to CIC University libraries. *Inside Google Books*. <https://web.archive.org/web/20110204173021/http://booksearch.blogspot.com/> (last checked 22 Mar 2026).
- Barthold, W. [= Bartol'd, Vasilij V.]. 1899. Russische Arbeiten über Ostasien: Jahresbericht für 1898. *Mittheilungen des Seminars für Orientalische Sprachen an der Königl. Friedrich-Wilhelms-Universität* 2. 140–151.
- Blake, Robert J. 1995. El latín notarial de un escriba bilingüe o 'bigrafo' del XIII. In Pérez González, Maurilio (coord.), *I Congreso Nacional de Latín Medieval: Actas (León, 1–4 de diciembre de 1993)*, 463–468. León: Universidad de León.

- Bunčić, Daniel. 2006. Systematisiertes Schriftenverzeichnis von Helmut Keipert. In Bunčić, Daniel & Trunte, Nikolaos (eds.): *Iter philologicum: Festschrift für Helmut Keipert zum 65. Geburtstag*, 21–47. München: Sagner.
- [Bunčić, Daniel. 2016. *Biscriptality: A sociolinguistic typology*. Edited by Daniel Bunčić & Sandra L. Lippert & Achim Rabus. With contributions by Anastasia Antipova, Carmen Brandt, Ekaterina Kislova, Henning Klöter, Alexandra von Lieven, Sandra L. Lippert, Helma Pasch, Achim Rabus, Jürgen Spitzmüller, and Constanze Weth. Heidelberg: Winter. — D. B. 2026.]
- Clarke, Cuthbert. 1777. *The true theory and practice of husbandry: Deduced from philosophical researches, and experience*, Newcastle: T. Slack.
- Consani, Carlo. 1988. Bilinguismo, diglossia e digrafia nella Grecia antica I: Considerazioni sulle iscrizioni bilingui di Cipro. In Campanile, Enrico & Cardona, Giorgio Raimondo & Lazzeroni, Romano (eds.), *Bilinguismo e biculturalismo nel mondo antico: Atti del Colloquio Interdisciplinare tenuto a Pisa il 28 e 29 settembre 1987*, 35–60. Pisa: Giardini.
- Consani, Carlo. 1989. Bilinguismo, diglossia e digrafia nella Grecia antica II: Le lettere di Filippo v e i decreti di Larissa (Schwyzer, *DGEEP*, 590). *Annali del Dipartimento di Studi del Mondo Classico e del Mediterraneo Antico* 11. 137–159.
- Consani, Carlo. 1990. Bilinguismo, diglossia e digrafia nella Grecia antica III: Le iscrizioni digrafe cipriote. *Orientamenti linguistici* 25. 63–79.
- Dale, Ian R. H. 1980. Digraphia. *International Journal of the Sociology of Language* 26. 5–13.
- DeFrancis, John. 1984. Digraphia. *Word* 35. 59–66.
- Dubrovin, Nikolaj F. (ed.). 1889. *Prisoedinenie Kryma k Rossii: Reskripty, pis'ma, reljaciï i donesenija*. Vol. 4 (1781–1782). Sanktpeterburg: Glazunov.
- Fell, Charles. 1729. *The lives of saints, collected from authentick records of church history*. Vol. 3. London: T. Osborne.
- Ferguson, Charles A. 1959. Diglossia. *Word* 15. 325–340.
- Gebhardt, Karl. 1974. *Das okzitanische Lehngut im Französischen*. Frankfurt am Main: Peter Lang.
- Grivelet, Stéphane. 2001. Introduction. *International Journal of the Sociology of Language* 150. 1–10.
- Gröschel, Bernhard. 2009. *Das Serbokroatische zwischen Linguistik und Politik: Mit einer Bibliographie zum postjugoslavischen Sprachenstreit*, München: Lincom.
- Jaquith, James R. 1976. Digraphia in advertising: The public as guinea pig. *Visible Language* 10(4). 295–308.
- Jaspers, Gerardus Johannes. 1979. Einleitung. In Stephan von Landskron, *Die Hymelstrasz: Mit einer Einleitung und vergleichenden Betrachtungen zum Sprachgebrauch in den Frühdruck (Augsburg 1484, 1501 und 1510)*. Amsterdam: Rodopi.
- Jespersen, Otto. 1904. *Phonetische Grundfragen*. Leipzig: Teubner.
- Joseph, John E. 1987. *Eloquence and power: The rise of language standards and standard languages*. New York: Blackwell.
- Keipert, Helmut. 1998. Überlegungen zur Terminologie der sog. nationalen Wiedergeburt der Slaven. In Bjørnflaten, Jan Ivar & Kjetsaa, Geir & Mathiassen, Terje (eds.), *A centenary of Slavic Studies in Norway: The Olaf Broch Symposium, Oslo 12–14 September 1996: Papers*, 121–141. Oslo: Norwegian Academy of Science and Letters.

- Keipert, Helmut. 2001. Erfundene Vergangenheit? *Jahrbücher für Geschichte Osteuropa* 49(2). 264–267.
- Keipert, Helmut. 2006a. *Glasnost'*: Zu den lexikographischen Voraussetzungen für begriffsgeschichtliche Untersuchungen im Russischen. In Thiergen, Peter (ed.), *Russische Begriffsgeschichte in der Neuzeit: Beiträge zu einem Forschungsdesiderat*, 1–21. Köln: Böhlau.
- Keipert, Helmut. 2006b. *Cerkovnoslavjanskij* in verordnetem Vergessen. In Jesenšek, Marko & Zorko, Zinka (eds.), *Jezikovna predanost: Akademiku prof. dr. Jožetu Toporišiču ob 80-letnici*, 129–145. Maribor, Ljubljana: Slavistično društvo, SAZU.
- Keipert, Helmut. 2006c. Über das Vergessen von Wissensbeständen in der Slavischen Philologie oder Jan Kollárs Abhandlung „Über die literarische Wechselseitigkeit“ in der „Danica Ilirska“ von 1836. *International Journal of Slavic Linguistics and Poetics* 44/45. 261–278
- Keipert, Helmut. 2008. Ljudevit Gajs „Zauberwörtchen“ *narodnost*. In Di Salvo, Maria & Moracci, Giovanna & Siedina, Giovanna (eds.), *Nel mondo degli Slavi: Incontri e dialoghi tra culture; Studi in onore di G. Brogi Bercoff*, vol. 1, 314–326. Firenze: Firenze University Press.
- Lafont, Robert. 1971. Un problème de culpabilité sociologique: la diglossie franco-occitane. *Langue française* 9(1). 93–99.
- Levašov, Pavel A. 1790. *Plěn i stradanie Rossijan u Turkov ili Obstožatel'noe opisanie bédstvennyx priključenij preterpěnnnyx imi v Car-gradě po ob'javlenii vojny i pri vojskě, za kotorym vlačili ix v svoix poxodax; s prioščeniem dnevnnyx zapisok o voinskix ix dějstvijax v prošedšuju vojnu, i mnogix strannyx, rědkix i ljubopytnnyx proizšestvij*. Sanktpeterburg: Bogdanovič.
- Ol'denburg, Sergej F. 1899. Otryvki kašgarskix i sanskritskix rukopisej iz sobranija N. F. Petrovskago, II. *Zapiski Vostočnago otdělenija Imperatorskago ruskago arxeologičeskago obščestva* 11 (1897–1898). 207–264.
- Pfeifer, Wolfgang (ed.). 2003. *Etymologisches Wörterbuch des Deutschen*. München: dtv.
- Pieridés, Demetrios. 1876. On a digraphic inscription found in Larnaca (read 6th April, 1875). *Transactions of the Society of Biblical Archaeology* 4(1). 38–43.
- Pierquin de Gembloux, Claude-Charles. 1840. *Histoire monétaire et philologique du Berry*. Vol. 1. Bourges: Ménage.
- Polivanov, Evgenij D. 1927. O literaturnom (standartnom) jazyke sovremennosti. *Rodnoj jazyk v škole* 1. 225–234.
- Robb, Thomas. 2003. Google as a quick 'n dirty corpus tool. *TESL-EJ* [= Teaching English as a Second Language—Electronic Journal] 7(2). <https://tesl-ej.org/ej26/int.html> (last checked 22 Mar 2026).
- Schneidewein, Johannes. 1740. *In quatuor institutionum imperialium D. Iustiniani libros commentarii*. Colonia Agrippina: Metternich. [1st edn. Argentoratum [Strasbourg]: Rihelius 1571.]
- Siebs, Theodor. 1897a. Review: van Helten, W. L. Zur Lexikologie [sic; spelled “lexicologie” in the reviewed book] des Altwestfriesischen. [...] Amsterdam [...] 1896. *Literaturblatt für germanische und romanische Philologie* 18(7). 219–223.
- Siebs, Theodor. 1897b. Review: Friesch woordenboek, bewerkt door Waling Dijkstra en dr. F. Buitenrust Hettena, [...] Leeuwarden 1896. *Zeitschrift für deutsche Philologie* 29. 552–557.

- Siebs, Theodor. 1957. *Deutsche Hochsprache: Bühnenaussprache*. 16th edn. (Ed. by Helmut de Boor and Paul Diels.) Berlin: De Gruyter. [1st edn. 1898.]
- Siméon, Rémi. 1889. Introduction. In idem (ed.), *Annales de Domingo Francisco de San Anton-Muñon Chimalpahin Quauhtlehuanitzin: Sixième et septième relations (1258-1612)*, V–XLIV. Paris: Maisonneuve et Leclerc.
- Tabak, Josip. 1950. Kako Vinaver prevodi Andersena. *Republika* 5. 316–321.
- Trediakovskij, Vasilij K. 1752. *Sočinenija i perevody kak stixami tak i prozoju*. Vol. 1, S.-Peterburg.
- Tyrwhitt, Thomas. 1782. *A vindication of the appendix to the poems, called Rowley's, in reply to the answers of the Dean of Exeter, Jacob Bryant, Esquire, and a third anonymous writer; with some further observations upon those poems, and an examination of the evidence which has been produced in support of their authenticity*, London: Payne.
- Unseth, Pete. 2008. The sociolinguistics of script choice: An introduction. *International Journal of the Sociology of Language* 192. 1–4.
- von Sallet, Alfred. 1875. Die Münzen der griechischen Könige von Salamis in Cypern und die denselben zugetheilten modernen Fälschungen. *Zeitschrift für Numismatik* 2. 130–137 + Tafel V.
- Zima, Petr. 1974. Digraphia: The case of Hausa. *Linguistics* 124. 57–69.