

UDK: 81'23/'24
DOI: 10.33669/KJ2024-35-06
Rad primljen: 12. 6. 2024.
Rad prihvaćen: 3. 11. 2024.

Pregledni naučni rad

Tomasz Kamusella

University of St Andrews
tdk2@st-andrews.ac.uk
<https://orcid.org/0000-0003-3484-8352>

Counting the World's Languages: The Politics and Discontents of Enumeration

Abstract: The concern with establishing the number of the world's extant languages appeared in Europe during the 18th century. In the 20th century, anthropological research drove this exercise, alongside the need to deal with an increase in publications produced in a variety of non-European languages across the decolonized states. Meanwhile, during the past two centuries, the counting of languages was underpinned with the millenarist program of translating the Bible into all the world's languages. The end of the Cold War heralded the rise of the internet. Only the languages that are officially 'counted' (enumerated and supplied with standardized codes) feature in cyberspace.

Keywords: concept of Einzelsprache, counting languages, cyberspace, dichotomy of dialect and language, internet, linguistic imperialism.

Introduction

The article grew out from an encyclopedia entry on the history of attempts to count the globe's languages for the forthcoming third edition of Elsevier's *Encyclopedia of Language and Linguistics* (Nesi & Milin, 2024). Reference works as a genre entail brevity, a presentation of the basic facts and crucial developments. Yet, in my research for this entry

a lot of background events were uncovered that shed much needed light on the story and its logic, which is steeped in power relations on the scale of the entire world. Most of the material could not be fitted into the aforementioned encyclopedia entry. Hence, instead of casting away what was gathered, I decided to present the story in the form of an article that allows for dealing with necessary detail and nuance.¹

The concern with counting and establishing the number of the world's extant languages is relatively recent. This idea appeared in Europe during the 18th century and as such amounts to a hallmark of (Western) modernity. In the 19th century, this preoccupation featured in the positivist (and supremacist) program of 'gathering all the knowledge.' In the following century, the leading goals that drove this counting exercise were, first, anthropological research, while after World War II, the need to come to terms with an unprecedented explosion in the printed material produced in a variety of non-European languages across the decolonized states. Meanwhile, during the past two centuries, the counting of languages was strongly underpinned with the Christian millenarist program of translating the Bible into all the world's languages (as a theological requirement for the 'second coming of Christ' [Russell, 1887]). The end of the Cold War largely coincided with the worldwide rise of the internet. Only the languages that are 'counted' (officially enumerated, internationally recognized, and supplied with machine-readable standard codes) feature in cyberspace. Otherwise, they are 'invisible' to software solutions that underpin the internet and IT equipment.

What is Counted?

(Modern) Humans (that is, *Homo sapiens sapiens*), as a species, developed the biological (evolutionary) capacity for speech (*Sprache* in German), which is commonly referred to in English with the uncountable noun 'language' that takes no plural form. This capacity is

¹ I thank Florian Coulmas for his kind invitation to write the aforesaid encyclopedia entry, and also for his advice and suggestions for improvement. A word of thanks also goes to the three anonymous reviewers. Obviously, I am responsible for any infelicities remaining.

actualized exclusively in (cohesive) human groups, leading to the rise of language (*Sprache*, speech) varieties, which in English are construed as languages (*Einzel Sprachen* in German) in plural. This English homonymy between 'language' (*Sprache*) and 'a language' (*Einzel Sprache*) masks the crucial difference between these two different phenomena. Hence, when it is necessary to make the difference obvious, the German-language specialist terms are employed for this purpose, namely, *Sprache* and *Einzel Sprache*.² The former is biological and as such part of nature, while the latter are of human making, hence, part of culture. Significantly, languages (*Einzel Sprachen*) are both the medium of man-made culture and a product of this very culture.

The Human species has only a single (type of) capacity for speech (*Sprache*), biologically wired into people's bodies through the process of evolution. If, for the purpose of counting languages, this capacity is identified as an *Einzel Sprache*, then there is just one language, 'Humanese,' in the (human) world, or rather the *ecumene* (evolutionary niche), that is, where people live. Such an intuition is visible in the Judeo-Christian (Middle Eastern and European) scholarly-cum-theological tradition of searching for the 'Adamic language.' This term denotes the language of the first person ('Adam,' Hebrew אָדָם *'adam*, meaning 'a man, person, human, Humanity') in the biblical version of the creation of Humanity (Eco, 1995: 95-99; Gen 2:19). In Antiquity, Jewish and Christian theologians identified Hebrew as this original language of all Humanity, mainly, because it was the language of the original of the Torah (Old Testament) (Augustinus, 1871: 121-124). In the biblical myth, the Judeo-Christian God punished Humanity for building the sky-high Tower of Babel. This punishment entailed the destruction of this tower and splintering (Hebrew בָּלַל *bālal* for 'to confuse') the Adamic language into numerous mutually incomprehensible *Einzel Sprachen* (Eco, 1995: 8-10, 16-18, 37-39; Gen 11:1-9).

² A glossary of specialist terms is appended at the article's end (prior to the references). For the sake of clarity, these terms are defined and cross-referenced as employed in this text.

At the anthropological level, speech variety is connected to the fact that humans live in relatively separate groups (Nettle, 1999: 60-114). The primary (evolutionary) function of language is social bonding among individuals that enables the construction and maintenance of stable and socially cohesive groups (Dunbar, 1992). Communication isolation entailed by the division of Humanity among such groups generates speech diversification (cf. Deutsch, 1966). Obviously, individual humans continually cross from one group to another lessening this isolation. But in the pre-modern world geographic barriers (for instance, mountain ranges, or large bodies of water) and sheer distance tended to spatially compound the social in its character separation of human groups. Less permeable and more long-lasting isolation results in greater language diversification ('linguistic distance') (Dixon, 1997: 35-36), in other words, in less mutual comprehensibility.

However, this incomprehensibility of speech varieties among spatially immobile and overlapping (through criss-crossing members) groups was not discrete in spatial or social terms. Speech difference was rather continuous, changing gradually from one group to another, from region to region. This trait of gradual change in speech among pre-modern populations gave rise to the linguistic concept of dialect continuum for analyzing it. Yet, gradualness can be measured but not counted (hence, in bucket may contain ten liters of water, but not *ten waters). For the sake of counting, languages must be imagined and implemented as discrete and self-contained entities ('billiard balls') in space, and as coterminous with a single society (speech community). Hence, first, the concept of such a countable and discrete language (Einzelsprache) must be developed.

It appears that writing was this crucial technology that allowed for the emergence of this concept of discrete Einzelsprache (γλώσσα *glóssa* in Greek, and *lingua* in Latin) in the Judeo-Greco-Roman world. In the ancient Middle East and the Mediterranean Basin, the tendency was to develop a different system of writing for each different Einzelsprache. Thus, around the first century CE, in this area, at that time coterminous with the Roman Empire, there were two graphically clear-cut

Einzel Sprachen that were widely acknowledged and employed in writing by this empire's elite, namely, Greek and Latin. Both written languages came with their own specific writing systems, to this day, eponymously known as the Greek and Latin (Roman) alphabets. In the fourth century, with the acceptance of Christianity as the empire's state religion, Hebrew jotted down in its own Hebrew letters was added as another written language of note to Greek as the empire's language of culture and Latin as the empire's language of politics and administration.

The vast empire's inhabitants and administrators clearly realized a high degree of speech difference among the population at large but did not see it as ('proper') languages. A speech variety without its own specific script and a definite role in the empire could not function as an Einzelsprache (cf. Adams, Janse & Swain, 2002). It was mere 'barbarian' gibberish or 'βαρ-βαρ *bar-bar*' in Greek. Noticing this power and technological differential between 'written languages,' and 'barbarian' (that is, unwritten) speech led, at the turn of the first millennium CE, to the coining of the Greek term διάλεκτος *diálektos* ('rural, regional speech') for the latter (yielding *dialectos* in Latin) (Kamusella 2016).

When the practise of writing and construing about linguistic variety led to the emergence of the concept of Einzelsprache at the beginning of the first millennium CE, there were just three 'true' (or acknowledged, written, enumerated, empowered) languages in the Roman Empire, namely, Greek, Hebrew, and Latin. The situation remained largely unchanged in the western half of this area, which following the great schism of 1054 corresponded to the ecclesiastical territory of the Catholic Church in (western and central) Europe. Latin was elevated there to the role of the leading official language of politics, religion, culture, and education, while the knowledge of Greek and Hebrew was cultivated for specialist purposes. In the 16th century, Protestantism and the Counter-Reformation encouraged the translation of the (Latin) Bible into some of the faithful's speech varieties, also known as 'vernacular' (meaning 'domestic, uncultivated') languages. This confessionally (ideologically) driven political process led to the rise of new Einzel Sprachen, each complete with its Latin-style grammar (Rosier-Catach 1988). Unlike earlier languages, Czech, French,

Hungarian, German, Italian, or Polish all share the same writing system of the Catholic Church's 'sacred language' of Latin. These varieties into which the Bible was not translated (or when such a translation was abandoned by the faithful) began to be referred to as 'dialects.'

Meanwhile, in 'Romania' (Ρωμανία), or the eastern half of the Roman Empire (so-called 'Byzantium'), a novel tradition of spreading Christianity ('evangelization' or 'Christianization') developed. Constantinople gave some neighbouring peoples across the empire's borderlands, or beyond in its sphere of influence, an acknowledgement in the form of a permit to translate the Greek-language Bible into a local prestige speech variety. As a result, this speech variety was elevated to the rank of another Einzelsprache among the empire's recognized languages. Each new Einzelsprache-in-making came complete with its own specific script (writing system). Gothic in the fourth century, Armenian, Georgian and Syriac in the fifth century, or (Church) Slavonic in the ninth century. As of the seventh century, the Arabs adopted the Judeo-Greco-Roman tradition of literacy for their new religion of Islam, steeped in the Arabic-language Quran and written in Arabic letters. The Islamic Empire and Muslim merchants spread this tradition from Northern Africa to what today is Indonesia and from Central Asia to the Sahel.

Literati's espousal of what now is seen as the 'normal' (that is, normative) dichotomy of language and dialect took place during the age of religious wars across Western and Central Europe. For a language to become a recognized Einzelsprache it had to be employed in administration, for publishing and as a medium of education. In contrast, dialects remained unwritten or rarely written with no empowering functions, spoken by low status social strata and populations in a state's far-flung regions. A tendency developed to see languages as 'better' and 'civilized,' while dialects as 'inferior' than languages, 'coarse' and 'uncultivated.' Both terms began also to be ranked in the terms of the aforementioned dichotomy. Typically, there are always more dialects than recognized languages within a given territory. The former are perceived as 'belonging to' (subsumed in) a (polity's prestigious) language. With the rise of Europe's global empires during the 17th and 18th centuries,

this dichotomy was already deemed as 'scientific,' that is, normative, normal and prevailing (Van Rooy, 2020: 95-216). Colonial administrators spread this dichotomy, together with the concept of *Einzelprache*, worldwide to the obsolescence or suppression of non-Western concepts for construing about and dealing with speech variety alone, or at the juncture with the technology of writing (Errington, 2007: 109-113).

Christianization was the preferred ideology for justifying European imperialism. Whatever wrongs European colonizers perpetrated these were readily explained away by missionaries. Missionaries also acted as linguists-cum-anthropologists in the field. First, they strove to acquire what they already saw as a local ('native') non-European language. Second, they described (or rather constructed) it in dictionaries and grammars written in line with the Western concept of *Einzelprache* (Villalba, 2023), and with an eye of translating the Bible (or at least the New Testament) into such a newly standardized language (Errington, 2007: 39). The purpose was evangelization, effective administrative control, economic exploitation, and ultimately the destruction of the local ('heathen,' 'native') culture (cf. Dixon, 1997: 103-115; Stoll 1982). To facilitate this 'civilizing mission' of evangelization on a global scale, a non-denominational British and Foreign Bible Society was founded in 1804. By the 1970s, the society had published translations of the Bible (or its parts) into over 1,400 languages (Whitaker, 1971: 1097).

Religious strife in Western and Central Europe brought about the production of polyglot Bibles during the 16th and 17th centuries. Apart from the original's languages (Aramaic, Greek, and Hebrew), they featured translations into Greek, Latin and Syriac, and later also into Arabic, Gééz, and Persian (Hendricks, 1967). At the same time, the genre of polyglot dictionaries emerged for ideological (confessional) and practical (administrative) purposes. A typical reference of this type paired the languages recognized in Antiquity (Greek, Hebrew, and Latin) with newly codified vernacular languages (Croatian, Czech, English, Dutch, French, German, Hungarian, Italian, Ottoman Turkish, Portuguese, Serbian, Slovenian, or Spanish) (Considine, 2019: 241, 301, 447-449). Polyglot dictionaries, for the sake of proselytizing and administration

also appeared in the Islamic (Persianate) world in the 14th century and four centuries later in China (Considine, 2019: 180, 206). But, to our knowledge, they did not lead to a non-European project of counting the world's languages. Meanwhile, European colonialism harnessed polyglot dictionaries for the purpose of building maritime empires (Considine, 2019: 618-619, 624, 657, 672-673, 692).

Who is Counting Languages?

Apart from establishing what is counted when the 'total number' of the world's languages is discussed, it is equally important to reflect on who is counting and for what purpose. Both the concept of *Einzelssprache* and its actualizations (languages) are products of human imagination and effort. As such they belong to the social reality (culture), which humans generate through their use of speech (*Sprache*). This social reality is perceivable ('visible') only to its producers-cum-participants, because its elements (concepts, ideas, or stories as expressed through the medium of speech, as actualized in *Einzelssprachen*) are stored in the human neo-cortex. Hence, languages (*not* speech) are imperceptible to non-humans (for instance, hominids), or more broadly to self-conscious beings that neither use language (*Sprache*) for group-bonding nor share the concept of *Einzelssprache*. In the latter case, languages are also 'invisible' to humans who do not know the concept of *Einzelssprache*, although they are aware of speech difference that occurs among people and their groups. Furthermore, devices used by physicists for measuring or otherwise for probing into matter or energy (that is, material reality) do not allow for detecting languages. For example, *Einzelssprachen* do not possess any weight to be weighed, length to be measured, or radiation to be assessed.

In general, counting languages is counting group-held imaginings about speech. Different groups and perceivers, even if they share the concept of *Einzelssprache*, may imagine speech variety X differently. Some would see it as an *Einzelssprache*, others as a dialect, and yet another group may disregard or overlook this variety as of no import for this group's culture, politics, social relations, or economy. The main dichotomy in such group perception is generated by the epistemic tension that exists

between emic (in-group) and etic (outside observer's) perspectives. For instance, standard Arabic (فصحى *Fuṣḥā*), as steeped in the language of the Quran, is well over a millennium removed from vernacular Arabic (dialects). No community speaks this standard at home, so Arabic-speaking children need to acquire it at school. Yet, most Arabic-speakers see *Fuṣḥā* as their 'native language' or 'mother tongue,' and disregard their vernacular Arabic varieties as 'undignified speech,' which is not fit for written use. On the other hand, foreigners who want to learn Arabic soon discover that no one speaks the standard, so for day-to-day purposes they must acquire a most suitable for their needs vernacular, choosing from among over 30 Arabic dialects, which are as different from one another as Spanish from French. In this configuration the temporally removed *Fuṣḥā* can be compared to Latin. To further complicate the matters, the Maltese who see themselves as a nation separate from the Arabs, speak their own Maltese language and as Catholics write it in Latin letters. Maltese may even take offence if it is proposed that their language is part of Maghrebi Arabic but have no problems to communicate with Arabic-speakers either in Tunisia or Libya (Kamusella, 2017).

Emic (insider's) view	1 language (<i>Fuṣḥā</i> ; 30 Arabic vernaculars disregarded as 'dialects'; Maltese seen as unrelated to Arabic)
Linguist's view	2 languages (standard languages of <i>Fuṣḥā</i> and Maltese; 30 Arabic vernaculars disregarded as 'dialects')
Etic (outsider's, foreigner's) view	More than 30 languages (30 Arabic vernaculars, <i>Fuṣḥā</i> , Maltese)

Arabic language or languages?

In Arabs' and Maltese's eyes Arabic and Maltese are two different languages. A scholar of Semitic linguistics may see them as (standard) varieties of a single Arabic-Maltese (supra-)language or dialectal continuum. Yet, what a non-Arabic-speaking foreigner may perceive happens to be as many as three dozen Arabic languages. Likewise, the name of a language (linguonym or glottonym) also results in variegated perceptions of entities to be counted as languages. The exolinguonym

(name given to a language by foreigners) of Korean suggests a single Einzelsprache, despite the existence of the two Korean states. Yet, the endolinguonyms (names employed by the language's speakers) indicate two languages, namely, North Korea's communist Einzelsprache of 조선말 *Chosŏnmal* and South Korea's democratic one of 한국말 *Hangungmal*. The same situation is observed in the case of the official languages of communist China and the Republic of China (that is, Taiwan). Foreigners see both as the single Chinese language. But in Chinese the communist Einzelsprache of 普通話 *Pǔtōnghuà* bears a different name than the democratic one of 國語 *Guóyǔ* (Kamusella, 2021a: 174). Furthermore, the difference between the two 'Korean' languages and the two 'Chinese' languages is emphasized through script. In *Chosŏnmal* Chinese characters (letters, graphemes) are avoided, unlike in *Hangungmal*. In the second case, *Guóyǔ* is written in traditional Chinese characters, while *Pǔtōnghuà* in simplified (communist) ones. However, in cyberspace, this scriptal difference between the two 'Chinese' Einzelsprachen shows in the Google Translate service and in the Chinese Wikipedia, which offer the choice between 'Chinese (Simplified)' and 'Chinese (Traditional).'

Historically speaking, in the Holy Roman Empire, at the turn of the 16th century, apart from official Latin, around ten 'German' chancery languages were used in the regional courts. All of them were locally referred to as *Teutsch*. With the development of printing, the number of the German 'printing languages' (*Druckersprachen*), based on these aforementioned chancery idioms, was gradually reduced for commercial reasons (Szulc, 1999: 61-66). The concomitant political developments eventually left the area with two standard languages in the modern period, namely, Dutch and German, which still share (almost) the same endolinguonym, or *Duits* and *Deutsch*, respectively. The early modern form of the endolinguonym for 'Dutch' *Duytsch* (later *Dietsch*) shows that initially it was identical with that for German.

The widely fluctuating number of Slavic languages, nowadays employed in numerous polities from the Baltic to the Adriatic and from Germany through Russia, offers an equally instructive story. In the Middle Ages,

most spoke about all the Slavic written and vernacular varieties using the single linguonym *Slovi(e/a)nski* ‘Slav(on)ic.’ At the turn of the 19th century, founding scholars of Slavic studies distinguished four Slavic languages, namely, Bohemian (present-day Czech and Slovak), Illyrian (today’s South Slavic languages), Polish, and Russian (present-day Russian, Belarusian, and Ukrainian). These four Einzelsprachen were regularly used in publishing and administration. At this time, also the Pan-Slavic theory developed. Its proponents claimed that only a single Slavic language exists, whereas the four aforementioned Einzelsprachen constitute this language’s ‘literary dialects.’ The conceptual tension between proponents of fewer and more Slavic languages has driven numerous political and cultural developments among Slavic peoples and their polities to this day. For instance, one of the Kremlin’s ‘justifications’ for Russia’s ongoing war on Ukraine is the claim that ‘the Ukrainians as a nation do not exist,’ because their ‘language is a dialect of Russian.’ In 1914, eight Slavic languages were employed in official use (Bulgarian, Croatian, Czech, Polish, Slovenian, Russian, Serbian, and Ukrainian). A century later, in 2017, as many as 13 Slavic Einzelsprachen functioned in the role of state languages (Belarusian, Bosnian, Bulgarian, Croatian, Czech, Macedonian, Montenegrin, Polish, Serbian, Russian, Slovak, Slovenian, and Ukrainian). Yet, the total number almost doubled to 24 Einzelsprachen if non-state Slavic languages were included in the tally, or even trebled to 39 if contested Slavic micro-languages were also counted (Kamusella, 2021b: 10-11, 19, 60, 99).

From the etic perspective, it is apparent that Norway has one official language, or Norwegian (endolinguonym: *Norsk*). Yet, it comes in two different but officially equal varieties, *Bokmål* (‘Book language’) and *Nynorsk* (‘New Norwegian’), which are as different from each other as Italian from Spanish. At home and in a given commune Norwegians use one of these varieties but at school acquire literacy in both. Bokmål is almost identical with Danish but for some spelling differences to better reflect its Norwegianized pronunciation. On the other hand, Nynorsk is mutually incomprehensible with Danish, but rather closer to Faroese and Icelandic (Haugen, 1966a). In bookstores publications

are *not* sorted in line with the variety in which they were written. However, for composing a single book, typically, only a single variety is employed. This rather strict separation maintained by Norwegian-speakers necessitated the creation of two Norwegian Wikipedias, one in Bokmål and the other in Nynorsk. Foreigners become aware of these complications, only when they want to acquire Norwegian. This entails deciding whether to learn Bokmål, Nynorsk, or both. In turn, as a bonus acquiring Bokmål gives learners access to Danish, while Nynorsk to Faroese and Icelandic. What is more, mastering both varieties allows one a good degree of comprehension in Swedish (Haugen 1966b).

When Yugoslavia was founded after the Great War, Serbo-Croato-Slovenian was proclaimed as the new state's official language. In interwar practice, this language came in two geographically and functionally delimited varieties, Serbo-Croatian, and Slovenian. The former was employed across the entire country as the state language, while the latter was limited to northernmost Yugoslavia (or what today is) Slovenia for local uses there. Meanwhile, Serbo-Croatian came in three scriptal sub-varieties. Two were official, in Cyrillic and Latin letters, the former for Orthodox Serbs, whereas the latter for Catholic Croats. Simultaneously, Bosnia's Slavophone Muslims continued using the Arabic script for writing and publishing. After World War II, in federal Yugoslavia, Slovenian was recognized as an Einzelsprache. Furthermore, biscriptal Serbo-Croatian replaced interwar Yugoslavia's triscriptal Serbo-Croato-Slovenian as the country's state language. Outside observers often colloquially dubbed both as 'Yugoslavian.' Apart from Slovenian, also Macedonian was excluded from this Serbo-Croatian linguistic commonality, and elevated to the status of a separate language. Macedonian is mutually comprehensible with Bulgarian, but not with Serbo-Croatian. What is more, it was agreed that Serbo-Croatian came in four republican varieties, namely, Bosnian, Croatian, Montenegrin and Serbian (Brozović & Ivić, 1988).

Following the two-decade-long breakup of Yugoslavia, at the turn of the 21st century, these four sub-varieties were elevated to the role of state (national) languages in independent Bosnia, Croatia, Montenegro, and

Serbia (Čirgić, 2011; Greenberg 2004). Yet, the officially defunct bis-criptal language of Serbo-Croatian continues to be employed on the web, including a Wikipedia in this politically defunct Einzelsprache. Hundreds of thousands of self-declared Yugoslavs claim this language as their mother tongue, especially in North America (Srebotnjak 2016). Furthermore, Croatian shares the composite character of Norwegian. The Croatian language is construed as comprised of three varieties (*narječja* ‘written dialects’), namely, Čakavian, Kajkavian, and Štokavian. The last one is identical with standard Croatian in official use, while the former two are defined as historical and regional literary varieties (languages) of Croatian, nowadays in limited use (Hrvatska, 2020). Interestingly, Kajkavski is fully intelligible with Slovenian, which is not the case of standard Croatian (Dzino, 2010: 50). Last but not least, in 2005, the Štokavian-based (micro-)language of Bunjevac, written in Latin letters, was introduced as a school subject in Serbia’s autonomous region of Vojvodina. Meanwhile, with Zagreb’s support, the region’s Croats maintain that Bunjevac is yet another variety of the Croatian language (Vuković 2009).

1921	1 language (Serbo-Croato-Slovenian, so-called ‘Yugoslavian’)
1945	3 languages (Macedonian, Serbo-Croatian [so-called ‘Yugoslavian’], Slovenian)
1995 (Dayton Agreement)	5 languages (Bosnian, Croatian, Macedonian, Slovenian, Serbian)
2007 (Constitution of Montenegro)	6 languages (Bosnian, Croatian, Macedonian, Montenegrin, Serbian, Slovenian)
Internet	7 languages (Bosnian, Croatian, Macedonian, Montenegrin, Serbian, Serbo-Croatian [so-called ‘Yugoslavian’], Slovenian)
Including unofficial and regional varieties	10 languages (Bosnian, Bunjevac, Čakavian, Croatian [=Štokavian], Kajkavian, Macedonian, Montenegrin, Serbian, Serbo-Croatian [so-called ‘Yugoslavian’], Slovenian)

Fluctuating number of Yugoslavia’s (and post-Yugoslav) Slavic languages

History of Counting Languages

The rise of European (Western) imperialism provided the socio-political context of and an impetus for presumably the first-ever official project of assessing the world's linguistic variety. In the late 18th century, Russia – officially recognized as a European empire in 1721 at the conclusion of the Great Northern War – was eager to lead the way in this then 'emerging' field of inquiry. Funds were lavished on the project of identifying and counting the world's languages. In the case of Europe, it was counting this 'Christian' continent's Einzelsprachen as recognized and codified by their speakers. In most cases these European languages had already come complete with their grammars, dictionaries, and translations of the Bible. But this process of 'identifying' languages across the rest of the world amounted either to taking record of the already established Western imperial impositions in this regard across the extant colonies or to new (or planned) arbitrary impositions of pro-imperial imaginings in these areas which had not been colonized yet. What the colonized and not-yet-colonized populations may have thought of this exercise and these impositions was not taken into consideration.

Russia's monarch Catherine II (r. 1762-1796), who proposed this project of identifying and counting the globe's languages, entrusted its realization to the Prussian scholar and explorer in Russian employment Peter Simon Pallas (1741-1811) (Dixon, 2010: 275). The first edition of his comparative dictionary of 'all the world's languages and dialects' consisted of two volumes and covered Europe and Asia. The standardized word lists were provided for 200 languages, 51 from Europe and 149 from Asia (Pallas, 1786-1789). Soon afterward, the Austro-Serbian scholar and educationalist in Russian service Teodor Janković-Mirijevski (Fiodor Ivanovich Iankovich de Mirievo, 1741-1814) prepared a four-volume second edition of this reference, which also covered Africa and the Americas. In total, 279 languages were presented, 55 from Europe, 171 from Asia, 30 from Africa, and 23 from the Americas (Pallas & Iankovich, 1790-1791).

At the turn of the 19th century, the 'civilizing mission' as equated with Christianization was adopted in the function of the main ideological

justification for the continuing expansion of the European empires. Translating the Christian Bible into 'native' languages became a yardstick for both gauging the success of this imperialist project and for counting the world's Einzelsprachen. Prior to the modern period, the Bible (or its portions) was translated into about 25 languages in Antiquity and the Middle Ages. The Reformation-related movement for translating the Bible into (predominantly Europe's) vernaculars had resulted in 30 more translations by the turn of the 17th century. The growing ideological coupling between colonization and Christianization brought about 40 more translations during the 17th and 18th centuries. By the Congress of Vienna in 1815, the Bible (or its portions) had become available in almost a hundred Einzelsprachen (Gerner, 2017: 155). This number correlated then well with the world's fewer than 300 languages, as established by Pallas's research team.

In the age of high imperialism, between 1815 and 1914, the Bible (or its portions) was translated into almost half a thousand more languages, pushing the grand total to nearly 600 (Gerner, 2017: 157). Yet, prior to World War I, anthropological research and imperial censuses already assessed the number of Einzelsprachen in the colonies at around 6,000, or ten times more (Gerner, 2017: 159). The evangelical translation challenge increased exponentially. Meanwhile, in 1872, the 8th International Congress of Statistics, held in St. Petersburg, introduced to the West's censuses as standard a question about one's Einzelsprache, or 'mother tongue' (variously interpreted, as the first, family, or community language) (Labbé, 2019: 53-56; *Report*, 1875: 37, 41). In 1886, the 7th Oriental Congress, held in Vienna, recommended that the (colonial) government of India conduct a survey of all the British colony's languages, so these Einzelsprachen could be covered in censuses (Mahulkar, 1990: 90). The Linguistic Survey of India, conducted between 1903 and 1928, registered 872 Einzelsprachen. At the same time, British colonial administration sought to reduce ('rationalize') the number of languages in which some education and civil services would have to be provided. For instance, the 1921 Indian census recorded 188 languages and 49 dialects (Singh & Manoharan, 1993: 4).

During the 1920s, a similar exercise of counting languages was conducted in the ideologically atheist Soviet Union. The goal was to provide education and (local) administration in all the languages of the communist polity's inhabitants with an eye to spreading the 'gospel' of marxism-leninism more efficiently. Initially, over 170 Einzelsprachen were identified, but the Soviet bureaucrats soon 'rationalized' the number down to 112 languages, which further plummeted to 104 Einzelsprachen in 1970 (Hirsch, 2005: 134, 143, 322). The Soviet model was followed in communist China, Vietnam, Laos, and Ethiopia where, respectively, 54, 53, 47 and 85 Einzelsprachen were identified (*Lao*, 2006: 4; Lewis, 1983: 20; Masako, 2013; Mullaney, 2011: 54, 83). Meanwhile, during the interwar period, the number of languages in which the Bible (or its portions) was available had grown by a third to around 800 by the close of World War II (Gerner, 2018: 148).

At the international level of political practice, only two Western (European) Einzelsprachen (English and French) were adopted as working languages in the League of Nations. After 1945, in the United Nations (UN) this elevated status was accorded to six languages from Eurasia (Arabic, Chinese, English, French, Russian, and Spanish). However, the UN's most ambitious document, or the Universal Declaration of Human Rights (1948) had been translated into 298 languages by 1999 (New Record, 2016). Decolonization and the global character of the Cold War made all Humanity's languages into potential media of ideological struggle between East and West (cf. Kowalski, 2015). For the 7th World Congress of Anthropologists and Ethnologists, held in Moscow in 1964, Soviet scholars prepared the then most comprehensive atlas of the globe's ethnic groups (Bruk & Apenchenko, 1964). This atlas depicted over 900 ethnic groups, which in Soviet scholarship were identified with their separate Einzelsprachen. This work was translated into English and remained in use worldwide as the most authoritative reference of this type until the fall of communism (Telberg, 1965).

In 1934, Presbyterian minister William Cameron Townsend (1896-1982) founded an evangelical organization Summer Institute of Linguistics in Dallas, Texas. Its goal was to harness linguistics and other

'scientific' methods for identifying and describing (that is, creating and standardizing) Einzelsprachen into which the Bible still had to be translated. This ideologically driven both 'scientific' and applied approach was not different from that employed in the Soviet Union and other communist states. By that time, the Roman Catholic Church still had had at its disposal more nuanced information on the world's extant languages (Drexel & Wimpssinger, 1934; Drexel, 1954-1958). Eight years later, in 1942, Townsend established another missionary organization Wycliffe Bible Translators USA. This organization was to use the expertise and training provided by the Summer Institute for the sake of accelerating the project of translating the Bible into all the world's languages (Stoll 1982: 3-6). To gauge this task, in 1951, the Summer Institute began publishing a reference, titled *Ethnologue: Languages of the World*. This title clearly indicated an espousal of the Soviet-style equation of languages with ethnic groups.

As of the 1960s, in its subsequent editions, *Ethnologue's* ambit was to identify and provide basic information about all the world's languages, alongside establishing their overall number. Theologians drew on some scholarly efforts to define and survey the globe's languages, which were conducted with an eye to classifying rather than counting these languages (cf. Meillet & Cohen, 1924 & 1952; Voegelin & Voegelin 1966 & 1977). The eighth edition of *Ethnologue*, published in 1974, identified 6,800 Einzelsprachen (*Ethnologue*, 2023; Grimes, 1974). The effort of coupling linguists with missionary theologians paid off, and during the Cold War period, the Bible (and its portions) was translated into almost 1,000 more languages (Gerner, 2018: 148). A non-confessional project of describing and counting the world's languages on the basis of censuses and official statistics took much longer to complete (Kloss & McConnell, 1974-1998) and never became as user-friendly as *Ethnologue*. An epitome of this situation was Charles Zisa's (1970) reference, available only as a microfiche, in which he recorded 18,000 linguonims.

Probing into the issue of the number of the world's languages from a purely scholarly perspective, Polish researcher Alfred F. Majewicz

arrived at the same number of the globe's languages, like *Ethnologue* (Majewicz, 1989: 242). But uniquely, in his work (completed in 1984 [Majewicz, 1989: 6]), this researcher attempted to classify the covered languages in a multifaceted manner, using genetic, areal, and typological approaches. On top of that Majewicz provided samples of the mainly non-Latin alphabet-based writing systems for over 300 Einzelsprachen. He was one of the first to remark that scripts make languages, thus, also must be taken into consideration when identifying the world's extant Einzelsprachen. To a degree, German linguist Florian Coulmas shared Majewicz's intuition. Coincidentally, the former's survey of the globe's writing systems came off the press in the same year (Coulmas, 1989). Meanwhile, a survey of all the globe's languages employed – to a varying degree – in writing and publishing was also attempted (Kloss & McConnell, 1978-2003). But its findings failed to brush off onto research in general or public discourse, unlike the single-volume comprehensive encyclopedias of the globe's writing systems, published in Britain, Japan, and the United States (Coulmas, 1996; Daniels & Bright, 1996; Sekai, 1993).

After World War II, decolonization and the Cold War fuelled a virtual explosion in publishing with the employment of numerous languages. Cataloging books and periodicals according to their subjects and names of authors turned out to be insufficient. Languages (and scripts) of this avalanche of printed material had to be also identified to make such publications accessible to readers. What is more, finding titles of interest among tens of millions required automation. In the early 1960s, the Library of Congress in Washington DC developed a MARC (MACHINE Readable Cataloging) standard for computer storage and retrieval of bibliographic data, including information on languages. To the same end, in 1967, the International Organization for Standardization proposed an ISO 639 standard for identifying Einzelsprachen. The running of this standard was entrusted to Infoterm (International Information Centre for Terminology), which UNESCO founded in 1971 in Vienna. The MARC list of languages provided three-letter codes that covered over 400 Einzelsprachen, while ISO 639 two-letter codes were extended

to almost 200 languages (Kamusella, 2012: 63-64). The ISO 639 standard covered these languages in which relatively developed publishing industries existed. The MARC list covered the very same languages, together with 200 more upcoming ones in which some publishing had already been happening.

Curiously, despite the Soviet Union's highly multilingual character (Vinogradov 1966-1968), the stark politicization of the languages in this communist polity, and the Kremlin's ambition to spread the ideology of communism around the globe, Soviet scholars were rather late to join the competition to classify and count the world's languages (Iartseva, 1982; Iartseva & Serebrennikov, 1980-1982; Serdiuchenko & Konrad (1976-1993). Yet, they were leaders in research on the globe's ethnic groups (cf. Bruk & Apenchenko, 1964), which gave Soviet academia an initial head start in this field. Perhaps, an explanation of this curious 'negligence' is the Western embargo on exports of computers and IT technology to the Soviet bloc (McCarthy, 1990). Soviet librarians and bureaucrats continued to rely on pen and paper, which certainly prevented them from tackling single-handedly thousands of bibliographic items in hundreds of languages. The economic stagnation of the late Soviet period, followed by the near-collapse of postcommunist Russia's economy did not encourage novel research, either. Initially, Moscow fell back on non-US-led research devoted to the world's languages and its number (Kloss & McConnell, 1978-2003), before developing during the past three decades, a similarly traditional in its approach project of describing and counting the globe's languages already after the fall of communism (*Iazyki* 1993-).

The automation and computerization of data processing, spearheaded by the Library of Congress and the UN, became a basis for the rise of the internet at the turn of the 1990s. This moment coincided with the end of the Cold War, allowing for the worldwide spread of the fledgling internet and the aforementioned cataloging standards. Initially, only English was employed on the web, but the rapid globalization of the world's economy and cyberspace entailed an accelerating increase in the use of multiple languages and scripts. The European Union (EU)

and India were at the forefront of these changes. The former ensures the translation of all official business into the 24 official languages of its member states (Languages, 2024a). On the other hand, as of 2011, the 8th Schedule to India's constitution lists 22 official languages in which administration and education are provided at the level of the country's federal states (Constitutional, 2021). The EU's official languages are written in three scripts (Cyrillic, Greek, and Latin), while India's in as many as 13 different writing systems (Arabic, Bengali-Assamese, Devanagari, Gujarati, Gurmukhi, Kannada, Latin, Malayalam, Meitei, Odia, Ol Chiki, Tamil, and Telugu) (Osada & Onishi, 2017: 12-13).

The exponential increase in the production of texts online and in print (courtesy of desktop publishing) after the fall of communism and the breakup of the Soviet Union required a further nuancing in the standards employed for registering Einzelsprachen. The original ISO 639 standard was renamed as ISO 639-1 and released in 2002. Meanwhile, the MARC language list became a basis for developing the ISO 639-2 standard, which the Library of Congress released in 1998 (Development, 2010). Yet, all these moves amounted to streamlining the code lists already developed during the Cold War. The number of languages covered under ISO 639-2 standard grew by a quarter, reaching almost half a thousand. But the challenge was much greater. The appearance of many other languages, not covered by these two standards, but increasingly employed for the computer- and internet-enabled production of texts necessitated the development of a more capacious standard that would cover all the world's Einzelsprachen, namely, ISO 639-3. Its development was entrusted to the Library of Congress, which in turn drew on *Ethnologue's* list of languages and their three-letter codes, as originally developed by the Summer Institute of Linguistics. This new standard was published in 2007 and its maintenance was handed over to the Summer Institute. The Summer Institute was renamed with its acronym as SIL International to conceal this organization's evangelical and missionary origin, which may not be of liking to non-Christians (Kamusella, 2012: 72). At present, the ISO 639-3 list includes about 8,000 languages (*Guidelines*, 2023: 1).

Before writing	1 language (Humanese = Sprache, or evolutionary capacity for speech = 'Adamic language')
Antiquity	2 languages (Greek & Latin)
Fourth century	3 languages (Greek, Hebrew & Latin)
Ninth century	9 languages (Arabic, Armenian, Georgian, Gothic, Greek, Hebrew, Latin, Slavonic & Syriac)
15 th century	25 languages (Bible translations)
Early 17 th century	55 languages (Bible translations)
18 th century	95 languages (Bible translations)
1791	279 languages (Pallas & Iankovich, 1790-1791)
1914	600 languages (Bible translations)
1964	900 languages (Bruk & Apenchenko, 1964)
1974	6,800 languages (<i>Ethnologue</i>)
2000	23,000 languages (Linguasphere)
2022	26,000 'languages' (Glottolog)

Counting languages in the West: A Historical Overview

At the turn of the 21st century, the ambition to take a comprehensive etic snapshot of all the world's language varieties was whetted by the publication of David Dalby's massive two-volume reference, titled *The Linguasphere Register of the World's Languages and Speech Communities*, which covers around 70,000 linguonyms and ethnonyms, alongside 23,000 languages, dialects, language varieties and speech communities, including information on writing systems and geographical location. The mass of data is tagged and cross-referenced (Dalby 1999-2000). The project was too unwieldy in a published form so soon the data migrated online, where it is maintained in the dedicated Linguasphere Register of the World's Languages and Speech Communities (Linguasphere, 2023). Meanwhile, work began on a new standard ISO 639-6, which with the use of four-letter codes would register all the world's 'language variants' (that is, languages, dialects, language varieties and the like).

This task and the standard's maintenance were entrusted to company GeoLang that was incorporated in London in 2006. Three years later, this new standard was released. Yet, the level of detail and the mass of information proved too much for any practical, ideological, research, or business uses. Hence, this standard was withdrawn in 2014 (ISO 639-6, 2009; Withdrawal, 2014).

Meanwhile, scholars and more inquisitive users often expressed frustration with the confessional origins of SIL International's ISO 639-3 standard and with the lack of references in *Ethnologue* that underpins this standard. As a result, linguists Harald Hammarström and Sebastian Nordhoff decided to ameliorate this situation. In 2011, they launched Glottolog, or an open access online bibliographic database of the world's languages. During the past decade and a half, Robert Forkel, Martin Haspelmath, and Sebastian Bank also contributed to the project (About Glottolog, 2024; Welcome, 2024). In 2023, Glottolog registered 7,654 L1 ('native') languages (About Languoids, 2024), extended unique alphanumeric codes (four letters + four numbers) to 8,595 languages in total (Languages, 2024b), alongside 12,796 dialects, and 4,571 language 'families' (groups) (Forkel & Hammarström, 2022). Glottolog also furnished 425,000 bibliographic items on the covered languages, dialects, and language groups (References, 2024). In total, almost 26,000 unique 'glottocodes' were issued (Forkel & Hammarström, 2022).

Personal computers became more widely available at the turn of the 1990s, which coincided with the beginning of the internet. The medium of both is the technology of writing that comes in the form of scripts, composed of standardized (delimited) sets of letters (fonts, graphemes). The main electronic companies behind these IT developments were acutely aware of the need to standardize the display of fonts in their products because the latter needed to be interoperable for the internet to work properly. To this end, in 1991, the companies founded a Unicode Consortium responsible for developing and maintaining a standardized Universal Character Set for IT purposes. Meanwhile, specialists working for the International Organization for Standardization shared the same realization. In 1989, they proposed

an ISO/IEC 10646 standard for a Universal Coded Character Set. But to avoid harmful divergences, this standard was developed in parallel with Unicode's (*Information Technology*, 1990; ISO/IEC 10646, 2020; Kamusella 2012: 62-63). For the sake of developing language-specific keyboards and software, it became obvious that scripts must be made recognizable through standardized codes, like languages. In 1997, work commenced on an ISO 15924 standard of codes for the representation of names of scripts. Three years later this standard was published, and Unicode was entrusted with its maintenance (Early History 2023). At present (2024), the Universal Character Set is comprised of almost 150,000 characters, which are used for supporting 161 scripts and some 50 more sets of specialized graphic symbols (for instance, emojis) (Character, 2023; Codes, 2023; Supported, 2023).

The codes and standards are incorporated into the software architecture of the internet by the Internet Engineering Task Force (IETF), which also ensures their interoperability. The IETF began in 1986 as an informal governmental organization in the US, before moving under the supervision of the Internet Society (ISOC) in 1993. ISOC (founded in 1992) is a US nonprofit advocacy organization with its seats in in Reston, Virginia, and Geneva in Switzerland (Bradner, 1999). In 1995, the IETF developed the first-ever standard (RFC 1766) for internet language codes. It drew on the ISO 639 standard's two-letter language codes and the ISO 3166 standard's two-letter country codes (Alvestrand, 1995). In 2001, the standard was updated to include the ISO 639-2 standard's three-letter language codes (RFC 3066). In 2006, another update followed, including the incorporation of the ISO 15924 standard's four-letter codes for scripts, together with the UN M.49 standard's three-digit codes for geographical regions (that is, typically, states and dependencies) (Philips & Davies 2006). Three years later, another update (RFC 5646) was implemented with an eye to accommodate three-letter codes from the standards ISO 639-3 and 639-5, respectively, for 'all the world's languages' and for language groups ('families') (Philips & Davies 2009).

At present the IETF language tag (which is not case-sensitive) is composed of the following subtags: (1) a single primary language subtag, (2) up to three optional extended language subtags, (3) an optional script subtag, (4) an optional region subtag, (5) optional variant subtags (as introduced by the Internet Assigned Numbers Authority, IANA), (6) optional extension subtags, and (7) an optional private-use subtag. The syntax and uses of IETF language tags are described in BCP 47 (BCP 47, 2024). The acronym BCP stands for ‘Best Current Practice.’ BCPs, which at present number around 200, gather the obtaining standards and recommended approaches to defined issues on the internet and more broadly across IT. BCPs were introduced in 1995 (Rekhter 1995). In the framework of BCP 47, nowadays, 8,259 subtags (codes) are in use for languages (dialects and other varieties), 253 subtags (codes) for languages and varieties not covered by the ISO 639 standards (mainly, sign languages and regional varieties – ‘dialects’ – of Arabic), 222 subtags (codes) for scripts and their variants, 113 subtags (codes) for orthographic and scriptal variants of languages, 305 subtags (codes) for geographic regions (that is, mainly states and dependencies) (Exlang Codes, 2024; Geographic Region Codes, 2024; Language Codes, 2024; Script Codes, 2024; Variant Codes, 2024).

Hence, in 2024, it can be estimated that 8,625 languages (and their variants) are recognized and in registered (enumerated) use on the internet. The number is a sum of the subtags (codes) for languages, languages and varieties not featuring in ISO 639, and for orthographic and scriptal variants of languages.

Conclusion

Counting languages is imagining them. In the observed socio-political practice, the number these imaginings is corrected – downward or upward – by human groups that accept, change (usually split or merge), redefine, or reject such Einzelsprachen (and their writing systems) for widespread written and online use. At present, the Bible (or its portions) is available in almost 3,000 languages (Gerner, 2018: 146). Speakers of 724 languages have access to the entire Bible in their

(‘native’) Einzelsprachen, while speakers of further 1,617 languages can enjoy the full New Testament translated into their idioms (State 2022). UNESCO’s³ data base of book translations, Index Translationum, registers translations into and from over 700 languages (Original 2024). The Universal Declaration of Human Rights is now available in translation into 562 languages (About 2023). At present, Wikipedia is offered in 324 different Einzelsprachen (List of Wikipedias 2023), while the Google Translate service allows for translation among 132 languages, or in 133, if Chinese in traditional and simplified spelling are treated as two separate Einzelsprachen (Google 2024).

However, commercially viable book publishing happens in not more than 50 to 80 languages. This stark economic reality is exemplified by the availability of the most popular book series of the early 20th century – that is, J. K. Rowling’s *Harry Potter* – in translation into 85 Einzelsprachen (List of *Harry* 2023). Perhaps, the series’ popularity pushes for translations into less commercially viable Einzelsprachen. As a corrective, Index Translationum lists 50 ‘top’ languages into which books are most regularly translated and published (Top, 2024). Interestingly, although the worldwide shadow library z-lib offers publications in 185 languages, the drop-down list of languages, first, distinguishes the 31 main languages in which bulk of the material on offer is provided. Predictably, all the languages stem from Eurasia, but 16 are from Asia. This statistic shows the continuing dominance of Western (European) publishing, which now is increasingly curbed by publishers from Asia (Z-Library, 2024). Networking service X (formerly Twitter) offers its interface in 33 languages (including two scriptal – traditional and simplified – versions for Chinese) (Supported 2024). Over half of the internet content (still) comes exclusively in the medium of English. Languages in which at least half a percent of the internet content is available number a mere 24. Only nine of them stem from outside Europe (that is, Arabic, Chinese, Hebrew, Indonesian, Japanese, Korean, Persian, Turkish, Vietnamese) (Usage 2024).

³ UNESCO is an acronym that stands for the United Nations Educational, Scientific and Cultural Organization.

Glottolog	c 26,000
Linguasphere (GeoLang, ISO 639-6)	c 23,000 languages
Internet (BCP 47)	8,625
ISO 639-3	c 8,000 languages
Bible (or its portions)	c 3,000 languages
New Testament only	1,617 languages
entire Bible	724 languages
Index Translationum (all translations recorded)	Over 700 languages
Universal Declaration of Human Rights	562 languages
Wikipedia	324 languages
Google Translate	132 languages
<i>Harry Potter</i> (or global publishing industry)	85 languages
Index Translationum (most used target languages)	50 languages
Z-Library (main languages)	31 languages
X (formerly Twitter): Supported languages	33 languages
Languages in which at least 0.5% internet content is available	24 languages
Index Translationum (most used non-European target languages)	10 languages
United Nations (official languages)	6 languages
Non-Eurasian languages of <i>Harry Potter</i> translations	1 language

Counting the world's languages in 2023

Tellingly, most of the languages into which *Harry Potter* was translated, in which most internet content is available, and in which the bulk of the world's publishing takes place are Western (European) Einzelsprachen. All the languages into which *Harry Potter* was translated stem exclusively from Eurasia. The single exception is Maori, which functions as an official language in New Zealand. Likewise, almost no non-Eurasian

(that is, African, American, or Oceanian) Einzelsprachen are used in publishing, and precious little internet content is offered in them. As a rule of thumb, people outside Europe (Eurasia), if they wish to study at the university level, surf the web, and enjoy literature, they have no choice at present but to acquire a Western (European) Einzelsprache for these purposes. In non-Eurasian languages (including the majority of Asia's Einzelsprachen) only some school textbooks may be on offer, alongside a full or partial translation of the Bible. This gap in the use of languages for publishing and as media of university education between Europe (Eurasia) and the rest of the globe is a legacy of Western imperialism. But some positive changes can be observed, for example, facebook already offers its interface in 13 non-Eurasian languages (Amharic, Fula, Guarani, Haitian Creole, Hausa, Inuktitut, Inupiaq, Kinyarwanda, Malagasy, Shona, Somali, Swahili, and Tamazight [Berber]) (Select 2024). Likewise, the Google Translate service is available in at least 28 non-Eurasian languages (Amharic, Aymara, Bambara, Chichewa, Ewe, Guarani, Haitian Creole, Hausa, Hawaiian, Igbo Kinyarwanda, Krio, Luganda, Malagasy, Maori, Oromo, Quechua, Samoan, Sepedi, Sesotho, Shona, Somali, Swahili, Tigrinya, Tsonga, Twi, Yoruba, and Zulu) (Google 2024).

Decolonization ended the West's (Europe's) political and (to a lesser degree) economic domination over the world. But at present, the West's – or more correctly – Eurasia's cultural, educational, and linguistic imperialism continues largely unabated, also because it is rarely noticed. Most people (especially in the West), scholars and politicians included, consider the situation 'normal.' As a result, European (Eurasian) languages are more appreciated and empowered as Einzelsprachen than their non-Eurasian counterparts. The main cleavage in the global architecture of linguistic empowerment gapes between Eurasia and the rest of the world. Research, politics, and publishing for the entire globe of thousands of languages takes place mainly in a handful of European Einzelsprachen, with the addition of several Asian languages (Kamusella 2020). Hence, the counting of the world's languages is a highly ideologized exercise which shapes both human societies and

cyberspace. In essence this exercise constitutes a mirror placed in front of the face of the globe's unequal social, political, economic, and cultural relations. Mere enumeration without official acknowledgement and use in publishing, education, and on the web counts for little in the case of languages mentioned in a given reference or register. On top of that, the more cyberspace impacts Humanity, the more the world's languages exist as 'genuine Einzelsprachen' only if they happen to be in extensive use on the internet and as media of textual and audiovisual production and communication.

Glossary

Abjad – *see* **Script**.

Alphabet – *see* **Script**.

Concept of Einzelsprache – a Western (European) idea that **Sprache** should be construed as consisting of discrete units (like billiard balls), whose separateness is demarcated through **writing (systems)**, grammars and dictionaries. This specialist German-language term is used because in English the meanings denoted by it and **Sprache** are confusingly rendered with the use of the same lexeme (word), that is, 'language.'

Culture – *see* **Social reality**.

Dialect – a speech **variety** that does not enjoy a high political or social status. Often, a dialect, when 'genetically' (or otherwise) similar to a language (**Einzelsprache**) of importance in a given polity, is often posed as 'belonging to' this language.

Dichotomy of dialect and language – a Western (European) concept which evaluates speech **varieties** as 'better, civilized' languages (**Einzelsprachen**) and 'inferior, uncultivated' **dialects**. In this configuration dialects are seen as 'belonging to' languages, that is, as a **variety** of this or that widely recognized Einzelsprache.

Emic – an insider's (participant's, in-group) perspective. This adjective is derived from the linguistic term 'phonemic.' *See also* **Etic**.

Einzelsprache (*pl* Einzelsprachen) – *see* **Language** (countable noun).

Etic – an outsider's (foreigner's, out-group) perspective. This adjective is derived from the linguistic term 'phonetic.' *See also* **Emic**.

Language (countable noun, has a plural form) – an actualization of **Sprache** in a given human group (speech community), a language among many, or **Einzelssprache** in German. From the socio-political perspective, a language (Einzelssprache) is a speech **variety (dialect)** with an elevated political and social status, leading to its extensive use in **writing**, administration, publishing, and on the internet. Languages (Einzelssprachen) are the medium of and part of **culture** (that is, **social reality**).

Language (uncountable noun, no plural) – speech (**Sprache** in German), biological (evolutionary) capacity for speech. Sprache constitutes a part of nature (**material reality**).

Language variety – *see* **Variety**.

Material reality – all the universe's matter and energy that can be, for instance, weighed, measured, assessed or otherwise detected with instruments employed by physicists. *See also* **Social reality**.

Morphemic writing system – *see* **Script**.

Script – a **writing** system, or the use of the technology of writing for recording a specific **Einzelssprache**. Often a single script can be adapted for writing in (recording) numerous languages. This is the case of the Latin alphabet employed for writing English, Finnish, French, or Slovak. Scripts may 'map' **Sprache** at different linguistic levels. Alphabets map an Einzelssprache at the level of language sounds (phonemes). **Abjads** (consonantries) are a form of alphabets that shun mapping vowels. **Syllabaries** (as most often used in India) map an Einzelssprache at the level of syllables. **Morphemic writing systems** (as employed in China) map an Einzelssprache at the level of simple words (morphemes).

Social reality – in other words, **culture**, which humans generate through their use of speech (**Sprache**) for group-bonding (that is, social cohesion). All social reality is stored in the neocortex of individuals who share its elements and concepts, including **Einzelssprachen**. Social reality is 'visible' (that is, perceivable) only

to individuals who generate it and share its constitutive elements.

See also **Material reality**.

Sprache – *see* **Language** (uncountable noun).

Syllabary – *see* **Script**.

Variety – often qualified as ‘language, linguistic, speech’ variety. Any group-specific form of **Sprache**, which for political or other reasons is classified as a dialect or language (**Einzelsprache**).

Writing – a technology of the graphic recording of **Sprache** (speech).

Writing system – *see* **Script**. Yet, some scholars distinguish between both. They employ the term script for the visible form of a writing system (that is, letters jotted on paper), while writing system to refer to the structural make-up, namely, whether speech is recorded at the level of phonemes, syllables, or morphemes.

References

About Glottolog. 2024. <https://glottolog.org/about>.

About Languoids. 2024. <https://glottolog.org/glottolog/glottologinformation>.

About the Universal Declaration of Human Rights Translation Project (2023). Geneva: Office of the United Nations High Commissioner for Human Rights. <https://www.ohchr.org/en/human-rights/universal-declaration/universal-declaration-human-rights/about-universal-declaration-human-rights-translation-project>.

Adams, J N. Janse, M & Swain, S (eds.) (2002). *Bilingualism in Ancient Society: Language Contact and the Written Text*. Oxford: Oxford University Press.

Alvestrand, H (1995). RFC 1766: Tags for the Identification of Languages. Reston VA: IETF. <https://www.ietf.org/rfc/rfc1766.txt>.

Alvestrand, H (2001). RFC 3066: Tags for the Identification of Languages. Reston VA: IETF. <https://www.ietf.org/rfc/rfc3066.txt>.

Augustinus [=Aurelius Augustinus Hipponensis] (1871). *The City of God* (Vol. 2). Edinburgh: T. & T. Clark.

BCP 47. 2024. RFC Editor. <https://www.rfc-editor.org/info/bcp47>.

- Bradner, S (1999). The Internet Engineering Task Force. <https://web.archive.org/web/20141228181401/http://www.oreilly.com/openbook/open-sources/book/ietf.html>.
- Brozović, D & Ivić, P (1988). *Jezik, srpskohrvatski/hrvatskosrpski, hrvatski ili srpski*. Zagreb: Jugoslavenski leksikografski zavod „Miroslav Krleža”.
- Bruk, S I & Apenchenko, V S (1964). *Atlas narodov mira*. Moscow: Glavnoe upravlenie geodezii i kartografii Gosudarstvennogo geologicheskogo komiteta SSSR and Institut etnografii im. N. N. Miklukho-Maklaia Akademii Nauk SSSR.
- Character Counts: Unicode Statistics (2023). Mountain View CA: Unicode. https://www.unicode.org/versions/stats/charcountv15_1.html.
- Čirgić, A (2011). *Montenegrin Language in the Past and Present*. Podgorica: Matica Crnogorska.
- Codes for the Representation of Names of Scripts / Codes pour la représentation des noms d'écritures: ISO 15924 (2023). Mountain View CA: Unicode. <https://www.unicode.org/iso15924/iso15924-codes.html>.
- Considine, J (2019). *The Cambridge World History of Lexicography*. Cambridge: Cambridge University Press.
- Constitutional Provisions Relating to Eighth Schedule. 2021. New Delhi: Ministry of Home Affairs. https://www.mha.gov.in/sites/default/files/EighthSchedule_19052017.pdf.
- Coulmas, F (1989). *The Writing Systems of the World*. Oxford: Blackwell.
- Coulmas, F (1996). *The Blackwell Encyclopedia of Writing Systems*. Oxford: Blackwell.
- Dalby, D (1999-2000). *The Linguasphere Register of the World's Languages and Speech Communities* (2 vols.). Hebron, Wales: Linguasphere Press.
- Daniels, T D & Bright, W (1996). *The World's Writing Systems*. New York: Oxford University Press.
- Deutsch, K (1966). *Nationalism and Social Communication*. Cambridge MA: MIT Press.
- 'Development of ISO 639-2.' 2010. Washinton DC: Library of Congress. 18 Oct. <https://www.loc.gov/standards/iso639-2/develop.html>.

- Dixon, R M W (1997). *The Rise and Fall of Languages*. Cambridge: Cambridge University Press.
- Dixon, S (2010). *Catherine the Great*. London: Profile Books.
- Drexel, A & Wimpssinger, R (1934). *Atlas Linguisticus*. Innsbruck: Kifadruck & Kartographisches Institut.
- Drexel, A (1954-1958) *Sprachen der Erde* (3 vols.). Zürich: Akademie Verlag
- Dunbar, R I M (1992). Neocortex Size as a Constraint on Group Size in Primates. *Journal of Human Evolution* 22(6). 469-493.
- Dzino, D (2010). *Becoming Slav, Becoming Croat Identity Transformations in Post-Roman and Early Medieval Dalmatia* (Ser.: East Central and Eastern Europe in the Middle Ages, 450-1450). Leiden: Brill.
- Early History of ISO 15924. 2023. Mountain View CA: Unicode. <https://unicode.org/iso15924/develop.html>.
- Eco, U (1995). *The Search for the Perfect Language* [translated from the Italian by J Fentress]. Oxford: Blackwell.
- Errington, J (2007). *Linguistics in a Colonial World: A Story of Language, Meaning, and Power*. Oxford: Blackwell.
- ‘Ethnologue: History’ (2023). <https://www.ethnologue.com/history/>.
- Extlang Codes: BCP47 Language Subtag Lookup. 2024. <https://r12a.github.io/app-subtags/index.html>.
- Forkel, R & Hammarström, H (2022). Glottocodes: Identifiers Linking Families, Languages and Dialects to Comprehensive Reference Information. *Semantic Web* 13(6). 917-924. <https://content.iospress.com/articles/semantic-web/sw212843>.
- Geographic Regions Codes: BCP47 Language Subtag Lookup. 2024. <https://r12a.github.io/app-subtags/index.html>.
- Gerner, M (2017). ‘Why Worldwide Bible Translation Grows Exponentially’ *Journal of Religious History* 42(2). 145-180.
- Google Translate. 2024. <https://translate.google.co.uk/>.
- Greenberg R D (2004). *Language and Identity in the Balkans: Serbo-Croatian and Its Disintegration*. Oxford: Oxford University Press.

- Grimes, B F (1974). *Ethnologue: 8th Edition*. Huntington Beach CA: Wycliffe Bible Translators.
- Guidelines for the Use of ISO 639-3 Language Codes in MARC Records* (2023). Washington DC: Library of Congress. 12 Jan. <https://www.loc.gov/aba/pcc/scs/documents/ISO-639-3-guidelines.pdf>.
- Hanson, P (1992). *From Stagnation to Catastroika: Commentaries on the Soviet Economy, 1983-1991* (Ser.: Washington Papers, Vol. 155). New York: Praeger.
- Haugen, E (1966a). *Language Conflict and Language Planning: The Case of Modern Norwegian*. Cambridge MA: Harvard University Press.
- Haugen, E (1966b). Semicommunication: The Language Gap in Scandinavia. *Sociological Inquiry* 36(2). 280-297.
- Hendricks, D (1967). Profitless Printing: Publication of the Polyglots. *The Journal of Library History* 2(2). 98-116.
- Hirsch, F (2005). *Empire of Nations: Ethnographic Knowledge and the Making of the Soviet Union*. Ithaca, NY: Cornell University Press.
- Hrvatska narječja proglašena nematerijalnim kulturnim dobrom (2020). *Jezik. Časopis za kulturu hrvatskoga književnog jezika* 67(1). 1-4.
- Iartseva, V N (1982). *Iazyki i dialekty mira. Prospekt i slovník*. Moscow: Nauka.
- Iartseva, V N & Serebrennikov, B A (eds.). 1980-1982. *Teoreticheskie osnovy klassifikatsii iazykov mira* (2 vols.). Moscow: Nauka.
- Iazyki mira* (23 vols.) (1993-<2020>; book series). Moscow: Nauka, Indrik and Academia and St Petersburg: Nestor-Istoriia.
- Information Technology - Universal Coded Character Set (UCS) / Technologies de l'information - jeu universel de caractères codés* (1990). Geneva: International Organization for Standardization, Joint ISO IEC Technical Committee Information Technology.
- ISO 639-6:2009 Codes for the Representation of Names of Languages – Part 6: Alpha-4 Code for Comprehensive Coverage of Language Variants (2009). Geneva: International Organization for Standardization. <https://web.archive.org/web/20181031212034/https://www.iso.org/standard/43380.html>.

- ISO/IEC 10646:2020(en) Information technology — Universal coded character set (UCS). 2020. Geneva: International Organization for Standardization. <https://www.iso.org/obp/ui/es/#iso:std:iso-iec:10646:ed-6:v1:en>.
- Kamusella, T (2012). 'The Global Regime of Language Recognition.' *International Journal of the Sociology of Language* 218. 59-86.
- Kamusella, T (2016). 'The History of the Normative Opposition of Language versus Dialect: From Its Graeco-Latin Origin to Central Europe's Ethnolinguistic Nation-States.' *Colloquia Humanistica* 5. 164-188.
- Kamusella, T (2017). 'The Arabic Language: A Latin of Modernity.' *Journal of Nationalism, Memory & Language Politics* 11(2). 117-145
- Kamusella, T (2020). 'Global Language Politics: Eurasia versus the Rest.' *Journal of Nationalism, Memory and Language Politics* 14(2). 117-151 .
- Kamusella, T (2021a). *Words in Space and Time: A Historical Atlas of Language Politics in Modern Central Europe*. Budapest: CEU Press.
- Kamusella, T (2021b). *Politics and the Slavic Languages* (Ser.: Routledge Histories of Central and Eastern Europe). London; Routledge.
- Kloss, H & McConnell (eds.). 1974-1998. *Linguistic Composition of the Nations of the World / Composition linguistique des nations du monde* (6 vols.). Sainte-Foy, Quebec: Presses de l'Université Laval.
- Kloss, H & McConnell (eds.). 1978-2003. *The Written Languages of the World: A Survey of the Degree and Modes of Use / Les langues écrites du monde. Relevé du degré et des modes d'utilisation* (6 vols.). Sainte-Foy, Quebec: Presses de l'Université Laval. (NB: Vol. 6 is in Russian: *Pis'mennyye iazyki mira. Iazyki Rossiiskoy Federatsii. Sotsiolingvisticheskaia entsiklopediia*. Moscow: Academia.)
- Kowalski, M W (2015). *Antropolodzy na wojnie. O „brudnej” użyteczności nauk społecznych*. Warsaw: Wydawnictwa Uniwersytetu Warszawskiego.
- Labbé, M (2019). *La nationalité, une histoire de chiffres. Politique et statistiques en Europe Centrale (1848–1919)* (Ser.: Domaine histoire). Paris: Les Presses de Sciences Po.
- Language Codes: BCP47 Language Subtag Lookup. 2024. <https://r12a.github.io/app-subtags/index.html>.

- Languages. 2024a. Brussels: European Union. https://european-union.europa.eu/principles-countries-history/languages_en.
- Languages. 2024b. <https://glottolog.org/glottolog/language>.
- Lao People's Democratic Republic: Northern Region Sustainable Livelihoods Development Project* (2006). Vientiane: Prepared by the Government of Lao People's Democratic Republic for the Asian Development Bank (ADB). <http://www.asianlii.org/asia/other/ADBLPRes/2006/4.pdf>.
- Lewis, I M (1983). 'Introduction.' In Lewis, I M (ed.) *Nationalism and Self Determination in the Horn of Africa*. London: Ithaca Press. 1-22.
- Linguasphere Register of the World's Languages and Speech Communities (2023). <https://linguasphere.info>.
- List of *Harry Potter* Translations (2023). Wikipedia. 27 Dec. https://en.wikipedia.org/wiki/List_of_Harry_Potter_translations.
- List of Wikipedias by Language Group. 2023. Wikimedia: Meta-Wiki. 29 Mar. https://meta.wikimedia.org/wiki/List_of_Wikipedias_by_language_group.
- McCarthy, J (1990). *Terms for Soviet Access to Western Computer Technology*. Stanford CA: Hoover Press, Stanford University.
- Mahulkar, D D (1990). *Pre-Pāṇinian Linguistic Studies*. New Delhi: Northern Book Center.
- Majewicz, A F (1989). *Języki świata i ich klasyfikowanie*. Warsaw: Państwowe Wydawnictwo Naukowe.
- Masako, I (2013). *Politics of Ethnic Classification in Vietnam* (Ser.: Kyoto Area Studies on Asia, Vol. 23). Kyoto: Kyoto University Press and Melbourne: Trans Pacific Press.
- Meillet, A & Cohen, M (1924). *Les langues du monde* (Ser.: Collection linguistique, Vol. 16). Paris: Champion.
- Meillet, A & Cohen, M (1952). *Les langues du monde* (1 vol. & 1 atlas). Paris: CRNS.
- Mullaney, T (2011). *Coming to Terms with the Nation: Ethnic Classification in Modern China*. Berkeley CA: University of California Press.

- Nesi H & Milin, P (eds.) (2024) [Forthcoming]. *International Encyclopedia of Language and Linguistics* (3rd Edition). Amsterdam: Elsevier.
- Nettle, D (1999). *Language Diversity*. Oxford: Oxford University Press.
- New Record: Translations of Universal Declaration of Human Rights Pass 500 (2016). Geneva: Office of the United Nations High Commissioner for Human Rights. 2 Nov. <https://www.ohchr.org/en/ohchr/node/16711/new-record-translations-universal-declaration-human-rights-pass-500>.
- Original Language: Bibliographic Search: Index Translationum. 2024. Paris: UNESCO. <https://www.unesco.org/xtrans/bsform.aspx>.
- Osada, T & Onishi, M (eds.) (2017). *Language Atlas of South Asia* (Ser.: RIHN-Manohar Indus Project series). New Delhi, Daryaganj: Manohar.
- Pallas P S (1786-1789). *Linguarum totius orbis vocabularia comparativa* (2 vols). St. Petersburg: Iohannis Caroli Schnoor.
- Pallas P S & Iankovich de Mirievo, F I (=Janković-Mirijevski, T) (1790-1791). *Sravnitel'nyi slovar' vsiekh iazykov" i narechii* (4 vols). St. Petersburg: Breitkopf.
- Philips, A and Davies, M (2006). RFC 4646: Tags for Identifying Languages. Reston VA: IETF. <https://datatracker.ietf.org/doc/html/rfc4646>.
- Philips, A and Davies, M (2009). RFC 5646: Tags for Identifying Languages. Reston VA: IETF. <https://datatracker.ietf.org/doc/html/rfc5646.html>.
- References. 2024. <https://glottolog.org/langdoc>.
- Rekhter, Y (1995). Best Current Practices. Reston VA: IETF. <https://www.rfc-editor.org/rfc/rfc1818>.
- Report of the Delegates to the International Statistical Congress Held at St Petersburg in August, 1872*. 1875. Washington, DC: Government Printing Office.
- Rosier-Catach I (ed.) (1988). *L'héritage des grammairiens latins de l'Antiquité aux Lumières* (Ser: Bibliothèque de l'information grammaticale, Vol. 13). Paris: Société pour l'information grammaticale (distributed by Peeters, Louvain).
- Russell, J S (1887). *The Parousia: A Critical Inquiry into the New Testament Doctrine of Our Lord's Second Coming*. London: T. Fisher Unwin.

- Script Codes: BCP47 Language Subtag Lookup. 2024. <https://r12a.github.io/app-subtags/index.html>.
- Sekai no moji no zu Nori*. 1993. Tokyo: Yoshikawa Kobunkan.
- Select Language. 2024. facebook. <https://m.facebook.com/language/>.
- Serdiuchenko, G P & Konrad, N I (1976-1993). *Iazyki Azii i Afriki* (6 vols.). Moscow: Nauka.
- Singh, K S & Manoharan, S (1993). *Languages and Scripts* (Ser.: People of India: National Series, Vol. 9). New Delhi: Anthropological Survey of India and Oxford University Press.
- Srebotnjak, H (2016). Tracing the Decline of Yugoslav Identity: A Case for 'Invisible' Ethnic Cleansing. *Sprawy Narodowościowe* 48. 65-84.
- State of the Bible 2022 (2023). Oxford: Wycliffe Bible Translators.
- Stoll, D (1982). *Fishers of Men or Founders of Empire? The Wycliffe Bible Translators in Latin America*. London: Zed Press.
- Supported Languages and Browsers. 2024. X (formerly Twitter). <https://developer.twitter.com/en/docs/twitter-for-websites/supported-languages>.
- Supported Scripts: The Standard (2023). Mountain View CA: Unicode. <https://www.unicode.org/standard/supported.html>.
- Szulc, A (1999). *Odmiany narodowe języka niemieckiego. Geneza, rozwój, perspektywy* (Ser.: Rozprawy Wydziału Filologicznego). Cracow: Polska Akademia Umiejętności.
- Telberg, V (1965). *Telberg Translation to 'Atlas narodov mira.'* New York: Telberg Book Corporation.
- "Top 50" Target Languages: Index Translationum. 2024. Paris: UNESCO. <https://www.unesco.org/xtrans/bsstatexp.aspx?crit1L=4&nTyp=min&topN=50>.
- Usage Statistics of Content Languages for Websites. 2024. W³Techs: Web Technology Surveys. https://w3techs.com/technologies/overview/content_language.
- Van Rooy, R (2020). *Language or Dialect? The History of a Conceptual Pair*. Oxford: Oxford University Press.

- Variant Codes: BCP47 Language Subtag Lookup. 2024. <https://r12a.github.io/app-subtags/index.html>.
- Villalba A D (2023). 'A Contribution to the History of Missionary Grammars and Romance Languages Grammars: The Commensurability of Metalanguage and Categories in the Sixteenth Century.' In Savatovsky D *et al.* (eds.) *Language Learning and Teaching in Missionary and Colonial Contexts / L'apprentissage et l'enseignement des langues en contextes missionnaire et colonial*. Amsterdam: Amsterdam University Press. 143-166.
- Vinogradov, V V (1966-1968). *Iazyki narodov SSSR* (5 vols.). Moscow: Nauka.
- Voegelin, C F & Voegelin, F M (1966). *Index to Languages of the World*. Bloomington IN: Indiana University (2 vols., published as part of journal *Anthropological Linguistics* Vol. 8[6-7]).
- Voegelin, C F & Voegelin, F M (1966). *Index to Languages of the World*. New York: Elsevier
- Vuković, P (2009). »Bunjevački jezik« – korijeni, varijeteti, perspektive. Godišnjak za znanstvena istraživanja. 141-155. https://web.archive.org/web/20150402192905/http://godisnjak.zkvh.org.rs/attachments/article/242/godisnjak2009_139-155.pdf.
- Welcome to Glottolog 4.8 Comprehensive Reference Information for the World's Languages, Especially the Lesser Known Languages. 2024. <https://glottolog.org>.
- Whitaker, J (1971). *An Almanack for the Year of Our Lord 1972*. London: Whitaker.
- Withdrawal of ISO 639-6 (2014). Berlin: TC37/SC2 Meeting. 21 Oct. <https://web.archive.org/web/20190821224741/https://www.alvestrand.no/pipermail/ietf-languages/2014-October/012167.html>.
- Zisa, C A (1970). *Language Classification and Indexing with an Annotated Bibliography*. Washington DC: Language Information Network and Clearinghouse System, Center for Applied Linguistics.
- Z-Library: Select Languages. 2024. <https://z-library.se/>.

Brojanje jezika u svijetu: Politika i nezadovoljstvo u procesu enumeracije

Sažetak: Briga oko utvrđivanja broja postojećih svjetskih jezika pojavila se u Evropi tokom 18. vijeka. U 20. veku, antropološka istraživanja su pokrenula ovu vežbu, zajedno sa potrebom da se pozabavi povećanjem publikacija proizvedenih na različitim neevropskim jezicima **širo**m dekolonizovanih država. U međuvremenu, tokom posljednja dva stoljeća, brojanje jezika je bilo podržano milenističkim programom prevođenja Biblije na sve svjetske jezike. Kraj Hladnog rata najavio je uspon interneta. Samo jezici koji se službeno „broje“ (nabrojani i snabdjeveni standardiziranim kodovima) postoje u sajber prostoru.

Ključne riječi: brojanje jezika, dihotomija dijalekta i jezika, internet, koncept Einzelsprache, lingvistički imperijalizam, sajber prostor.