The Global Regime of Language Recognition

Tomasz Kamusella

The Global Regime of Language Recognition[1]

Tomasz Kamusella
University of St Andrews, Scotland UK

Abstract

There is no universally accepted working linguistic definition of a language; the distinction between a dialect and a language is a political question. On the basis of a discussion of this problem, the article proposes that the ISO 639 family of standards, issued by the International Organization for Standardization (ISO), amounts to the backbone of an emerging global regime of language recognition. This regime is being rapidly coaxed into being by the booming IT industry and by the internet, both of which require clear-cut and uniform standards on languages and their scripts in order to function efficiently and profitably. A potentially undesirable and divisive foundation of the regulatory regime, stemming from and meeting the distinctive sectoral purposes of the world of Evangelicalism and Bible translation, is a hurdle to be overcome in achieving a universally accepted system of language standards. Despite efforts by other actors, there is no viable secular alternative in prospect, because the religiously-grounded system has an established and substantial "first mover" advantage in the field.

Key terms: Bible translation, global regime, internet, ISO 639, language recognition, proselytism

1. Introduction: Global-wide uniformization

The history of what is dubbed "modernity" can be usefully equated with the spread of various technologies and cultural practices, first, from the Western European "core" to what became the "West" (or the "rich North," as today's early 21st-century parlance of politics and journalism has it), and then, to the rest of the inhabited world (or the "poor South"). The most renowned and distinctive examples, include the calculation of the successive (solar) years beginning with Year One of the "Common Era" (the contemporary preferred term for the more ethnoreligiously specific *Anno Domini*); the currently standard (Gregorian) calendar and its divisions (seven-day weeks and its 12 months); and the splitting of the day into 24 hours of equal duration, measured by ubiquitous clocks and watches that construct each hour composed of 60 minutes. In turn, every minute is comprised of 60 seconds. (Blackburn and Holford-Strevens 1999) Amongst other relevant examples it is worth enumerating the division of the world into 24 time zones, the counting of which starts from the 'Prime Meridian of the longitude 0°,' which, under an international

agreement in 1884, passes through the Royal Observatory located in the London borough of Greenwich (*International* 1884).

The uniformization[2] of the systems of units of metrology, proposed in revolutionary France, became accepted worldwide in the wake of the signing of the Metre Convention in 1875. This led eventually to the development of an intricate system of seven basic units and a growing number of "derived units," the SI system or the "International System of Units" (from the French ***S****ystème **i**nternational d'unités*) (Goldman and Bell 1986).

The push for organizing and regulating the movement of road traffic on a global scale seems to have emanated from Napoleon's French Empire. It secured an almost universal acceptance for right-hand traffic, the notable exceptions being many former British colonies, Indonesia and Japan. Subsequently, the Multilateral Road Traffic Convention signed at Geneva in 1949 allowed for the standardization of road signs and traffic rules (Kincaid 1986).

Technology, which, in the wake of the industrial revolution, became ubiquitous in the 19th and 20th centuries has required an exponentially growing degree of uniformization in the construction of the basic units of machinery (for instance, batteries, transistors, four-stroke engines), in the production of 'stuffs' on which they run (for example oil measured in barrels, 95 octane unleaded gasoline, 220 volt electric current), and of goods, products and services in which they are used (for instance, train or bus public transport, private cars, household mechanical devices). The Cold War confrontation, however, introduced a high degree of international dichotomy into this unprecedented uniformization, just as that process took off in earnest in the mid-20th century.

The Soviet Union developed its own, separate set of technical standards to be employed in its industry and in the day-to-day life of the polity's inhabitants. The system was dubbed GOST, an acronym derived from the Russian term *gosudarstvennyi standart* for 'state standard.' It emerged in 1925 and after World War II was accepted across (or rather forced on) the entire Soviet bloc. Following the subsequent breakups of the bloc and of the Soviet Union, the use of GOST became limited to the Commonwealth of Independent States (CIS), and today is administered by the Euro-Asian Council for Standardization, Metrology and Certification and Russia's Rosstandart (Federal Agency on Technical Regulating and Metrology). In a nutshell, GOST, originally designed for the whole world, survives as a regional system in the post-Soviet section of Eurasia (About Organization 2011; About Agency; *Klassifikator* 1978).

The last significant Cold War battle in the arena of technological standards was fought between the SECAM and PAL color television analog systems. SECAM, or *Séquentiel couleur à mémoire*, the French for "Sequential Color

with Memory," was developed jointly by French and Soviet engineers during the 1960s. At the same time German and British engineers came up with another system, Phase Alternating Line, abbreviated to PAL.[3] In the 1990s the former Soviet bloc states migrated from SECAM to PAL, alongside the former Soviet republics of Estonia, Georgia (in the early 2000s), Latvia, Lithuania and Ukraine. France, the co-leader of SECAM with the Soviet Union, ceased broadcasting in the system in 2011 (coinciding with the change from analog to digital television transmission). At present, SECAM is limited to a smaller group of CIS states than GOST (Crane 1979; Information 1997).

The rise of the currently accepted world-wide standards in technology is associated with the International Federation of the National Standardizing Associations (ISA), founded in 1927. It was dissolved in 1942, but reappeared in 1946 as the International Organization for Standardization with its headquarters in Geneva.[4] On the one hand, it functions as one of the many global organizations more or less formally connected to the United Nations, while on the other, during the Cold War it was the "West's reply" to the Soviet GOST system (Craig 1994: 196, 222; ISO 2011; Latimer 1997).

2. Languages and technologically-driven uniformization

In the contemporary world, rapidly growing world-wide commerce, transport, travel and communication continue to be facilitated by the explosion of information technology; it sustains the cyberspace of the internet and makes it available to billions of people through ubiquitous personal computers. This has required an unprecedented level of further homogenization of standards in an ever growing number of fields of technology and even of social life. The most visible merger of technology and social life, driven by the developments in technology to meet social needs and economic interests, is epitomized by the Unicode Standard, or the Universal Character Set.[5] The idea of Unicode appeared in 1987-1988, following discussions on the subject among Apple and Xerox engineers. In 1991, the two companies, together with the RLG, Sun Microsystems, Microsoft and NeXT founded a Unicode Consortium.[6] The ambition was to encode in a standard manner, for computer and other IT use, characters of all scripts employed nowadays (and in some cases, also in the past) for writing and publishing. In early 2011, the system covered 93 scripts with about 109,000 characters.

In 1989, the International Organization for Standardization proposed its own draft standard ISO / IEC[7] 10646 for a universal character set, and published the first draft of this set a year later. Competition between ISO / IEC 10646 and Unicode was unwelcome and unsustainable, and the two standards were unified in 1990, though some differences remain between them. In practice the Unicode Consortium is responsible for maintaining

the standard and its further development. In recognition of the consortium's achievements, in 2004, the International Organization for Standardization entrusted it with the new ISO 15924 standard for conferring letter and numeric codes on writing systems (ISO 15924 2011; Summary 2011; The Unicode Standard 2011; UTF-8 2011).

The perceived normalcy of the fact that people across the world can write and communicate in their languages in different scripts via the internet, computers and a wide variety of electronic devices with no clashes of incompatible systems (quite a frequent occurrence during the 1990s), is possible only because of the success of the coordinated Unicode – ISO / IEC 10646 and the ISO 15924 standards. At present practically all 600-odd languages employed in writing and in publications, along with their 160-odd scripts, are covered by Unicode. One can write in the covered languages with the use of their specific scripts, because software companies, drawing on the standardized universal character set, developed similarly standard keyboard layouts for these languages and their specific scripts (Codes 2011; Languages 2011).

Earlier efforts to come up with standardized lists of language codes were connected to the task of sorting and ordering a ballooning quantity of information produced in paper form between the end of World War II and the onset of the internet epoch in the 1990s (Price 1963). At the forefront of this endeavor was the United States' Library of Congress with its aspiration to obtain and classify the entire (relevant) publication output of the world. In the 1950s, the reliance on paper and manual systems of cataloging and managing information on the library's holdings proved unsustainable. Hence, in the early 1960s, the MARC (**MA**chine **R**eadable **C**ataloging) standard was developed for computer storage and retrieval of bibliographic data (Intner 2007).

Among other MARC bibliographic elements, the Library of Congress came up with a MARC list of codes for languages in which the library's holdings were published. To the same end, in 1967 the International Organization for Standardization proposed the ISO 639 standard (ISO/R 639 2011). With its two-letter codes it covers the globe's 184 main languages, but the growth of book and periodical production in many other languages, spurred on by desktop publishing, especially in decolonized states, showed that the number of codes available in the ISO 639 standard was not enough for cataloging publications in languages other than those represented on the list. Hence, in 1998 the ISO 639 standard was split into separate component parts, in line with ISO standard procedures. The changed name of ISO 639-1 was allocated to the existing standard, while a new, more capacious, standard of three-letter codes, dubbed ISO 639-2, was entrusted to the Library of Congress for maintenance. This move on the part of the International Organization for Standardization amounted to the recognition of the success of the Library of Congress's MARC codes for languages. The ISO 639-2 list, though not identical with the MARC list of

language codes, is in practice based on the latter. In 2007 both lists, ISO 639-2 and MARC, contained 429 codes for languages and 55 for language groups (Codes for the Representation of Names of Languages 2011; ISO Language 2011; MARC 2007).

The ISO 639-1 standard[8] has become dormant, as a result of becoming in effect a subset of ISO 639-2. No additions to the list have been recorded since 2003. The conditions set out for such registration require
- a significant body of existing documents (specialized texts, such as college or university textbooks, technical documentation manuals, specialized journals, subject-field related books, etc.) written in specialized languages
- a number of existing terminologies in various subject fields (e.g. technical dictionaries, specialized glossaries, vocabularies, etc. in printed or electronic form)
- a recommendation and support of a specialized authority (such as a standards organization, governmental body, linguistic institution, or cultural organization)
- [a significant] the number of speakers of the language community
- the recognized status of the language in one or more countries
- [and] the support of the request by one or more official bodies (ISO 639 Joint 2000).

In short, it requires features common to languages used by millions of speakers in speech and writing, officially and colloquially, in all the domains of modern life. It would be difficult for a language to meet the requirements unless it was made an official language by a state.

The criteria for an ISO 639-2 are less onerous:
- the request for a new language code shall include evidence that one agency holds 50 different documents in the language or that five agencies hold a total of 50 different documents among them in the language. Documents include all forms of material and is not limited to text.
- If the criteria above are not met the language may be assigned a new or existing collective language code. The words *languages* or *other* as part of a language name indicates that a language code is a collective one (ISO 639 Joint 2000).

Other restrictions are as follow:
- a single language code is normally provided for a language even though the language is written in more than one script.
- a dialect of a language is usually represented by the same language code as that used for the language. If the language is assigned to a collective language code, the dialect is assigned to the same collective language code. The difference between dialects and languages will be decided on a case-by-case basis.
- a language using more than one orthography is not given multiple language codes (ISO 639 Joint 2000).

However, the ISO 639-2 standard became obsolete even faster than its predecessor, ISO 639 (that is, the current ISO 639-1). ISO 639-2, with its 460-odd codes at present, was long ago outstripped by Unicode, which supports over 600 languages in about 160 scripts. What is more, ISO 639-2 is mostly silent on the topic of writing systems, tacitly assuming

Romanization, or reducing other scripts to the Latin alphabet for the sake of cataloging and library science. It is an outdated approach. In the 1990s users of the internet, writing in languages that normally are not recorded in Latin characters, had no choice but to write phonetically in the Latin script, because of the limitations imposed by the available technology. But beginning at the turn of the 21st century, computer interfaces supporting scripts other than Latin have become widely available. This process of the "emancipation" of non-Latin scripts was completed in 2010 when the Internet Corporation for Assigned Names and Numbers began to approve a growing number of other scripts for internet addresses (ICANN 2010).

The internet and computer technology have become the basis and infrastructure of global-wide commerce, travel and information exchange. As a consequence, the commercial necessity to achieve economies of scale has entailed making this infrastructure accessible to as many speakers of different languages and users of various scripts as practical at a given moment. Unicode has been developed in close partnership with the largest software companies. Because these companies are commercially and existentially concerned with the realities of the internet-based economy, they may be more appropriate drivers of the development of an international technical standard relating to the internet than the Library of Congress, with its localized mandate to serve the needs of the US Congress and the American reader. The Library of Congress gathers and catalogs publications and information available in about 500 languages. Unicode furnishes software solutions for these languages and their scripts, but in addition also gives instruments to speakers of further 100-odd languages (as yet unrepresented in the Library of Congress) to produce publications and information in them, and in scripts specific to these languages. The tacit assumption is that speakers and users of the languages and scripts overlooked by the Library of Congress will purchase language-specific software solutions prepared by leading software companies and will thus become active participants in the global economy. In line with the theory of the "long tail," (Anderson 2006) speech communities of 20,000 to 50,000 literate people, living in a developed state in the "rich North," are big enough to justify investing in the development of such software solutions. This point was recently made by the development of various standard computer interfaces for the Kashubian language spoken by about 53,000 people in northern Poland (Kashubian Keyboard 2011; Wyniki 2008).

3. When is a language?

In 1926, the US linguist, Leonard Bloomfield, in line with the positivistic tradition attempting to refashion linguistics into an exact science, published his boldly entitled article "A set of postulates for the science of language" in the renowned journal *Language*. Among others, he gave a seemingly clear-cut definition of "language" (in the sense of "a language")

and "dialect." He proposed that languages are mutually incomprehensible language varieties, while dialects are mutually comprehensible language varieties. But putting this definition to a simple test quickly proves that it does not work. (Perhaps, among others, it is a reason why linguistics still has more to do with the social sciences than with physics or mathematics.)

The Slavic languages of Slovak and Czech are mutually comprehensible (Nábělková 2007), while the Romance standards of Moldovan and Romanian are, for all practical reasons, identical in speech and writing (Ciscel 2006). Low German, though comprehensible to speakers of Netherlandish (Dutch and Flemish, construed as a single language), the former is not a dialect of the latter but of the German language, though its speakers do not understand Low German (Gooskens 2006). Many dialects of Arabic are mutually incomprehensible but speakers of the Libyan dialect have no problems communicating with speakers of the Maltese language (Cf Durand 2009; Ferguson 1959; Zammit 2000). Likewise, dialects of Chinese can be as distinct vis-à-vis one another as Polish and Russian, or even as Russian and German (cf Künstler 2000). To further complicate the picture, Haugen (1966) showed that the phenomenon of mutual comprehensibility may be asymmetric (speakers of language A understand language B better than speakers of the latter comprehend the former), as in the case of certain pairs among the Scandinavian Germanic languages. Likewise, Portuguese-speakers tend to understand Spanish better than Spanish-speakers Portuguese (Jensen 1989). Hence, mutual comprehensibility is not a black-and-white phenomenon, but is characterized by shades of gray; a game of percentages, really.

The aforementioned examples indicate that there is something else at play other than the internal logic of language, when the decision is taken as to what is a language and what is a dialect. It is people – the sole producers and users of what we tend to refer to as "languages" – who take such decisions. A language is as much an artifact of culture as a teacup,[9] so the field falls out from the usual ambit of linguistics. As linguists say in their technical parlance, extralinguistic factors prevail in this case. The Jewish linguist of the Yiddish language, Max Weinreich, faced with this dilemma, quoted in his 1945 article a commonsensical answer to it, proposed by a student of his (whose name Weinreich did not record). In this view, a language is a dialect with an army and navy (*A shprakh iz a dialekt mit an armey un flot* in the Yiddish original). Ergo, while the proper subject of enquiry for linguistics is dialects, the study of languages in the plural and the differentiation among languages constitute an apt field of interdisciplinary study for social scientists who have mastered elements of linguistics, and vice versa, for linguists who do not shy away from the method of the social sciences. The latter are actually known as sociolinguists, their discipline having come into its own in the mid-20th century (Wodak, Johnstone and Kerswill 2011: 11-83).

One of the pioneers of sociolinguistics was the German linguist and specialist in the linguistics of minorities, Heinz Kloss. He managed to clothe the dilemma of how to distinguish between "dialect" and "language" in an analytical apparatus. He looked at the dialectal context of the emergence or rather creation of languages, and proposed the concepts of *Abstand* and *Ausbau* languages (Kloss 1967). The former are unambiguously mutually incomprehensible, or "stand away" from each other as the German term has it ("languages by distance"). Obvious examples may be German and Polish or Finnish and Swedish. Although the speakers of each of the two pairs of languages live side by side, their languages belong to different dialect continua. "Dialect continuum" denotes a contiguous geographic area in which dialects change from one to another (that is, from village to village, from town to town, from region to region) in a gradual manner, always retaining a fair degree of mutual comprehensibility. The border between two dialect continua also marks the line of discrete incomprehensibility between dialects spoken in neighboring villages or towns. Dialects from such adjoining dialect continua fashioned into languages are Abstand.

In contrast, Ausbau languages languages ("languages by extension, or construction") are developed within a single dialect continuum from mutually comprehensible dialects. The demarcating difference between such languages is consciously built or constructed as is implied in the German term. In this way the difference is deepened to the degree that is desired and considered appropriate by the speakers and fashioners of these languages. A typical end-product of such "difference construction" as pursued from the mid-19th to mid-20th centuries is the pair of mutually comprehensible languages, Czech and Slovak. At present we can observe the manufacturing of difference among Bosnian, Croatian, Montenegrin and Serbian, each of which coalesced as a language in its own right at the turn of the 21st century in the wake of the fragmentation of Yugoslavia, followed by the breakup of the Serbo-Croatian language (Greenberg 2004).

Dialects are the natural form of language; in other words they are as old as the modern human's capacity for language[10] (cf Laks 2008). The idea that there exist languages that are somehow "better" or "more developed" than dialects coincided with the invention of writing, a technology which is only five millennia old (Kuckenburg 2006: 151-152). The usually unarticulated received assumption draws the line between a language and a dialect, by pointing out that the former has the standing of a written form, while the latter is not employed in writing (cf Kuckenburg 2006: 131-149; Wicherkiewicz 2003). Written languages were the preserve of the narrow stratum of literati until the onset of modernity, which, among other consequences, entailed universal literacy guaranteed by free and compulsory popular elementary education. In this manner a language became an instrument of politics and mobilization on a par with an army or religion. Modern states tend to confer the status of official language on

a selected dialect (or a written language of earlier times). In turn the shape of this official language is uniformized (or standardized) through state-approved authoritative dictionaries and grammars that become the linguistic basis of school textbooks. The standard language is born. The educational system spreads its command among the entire populace of a state, which is subsequently reinforced by the constant and intensifying use of this language in state offices and mass media.

The process leads to a rarely perceived paradox of anachronism: scholars and the everyman tend to claim that a language consists of dialects, construed as its offspring. But, in the order of things, dialects are older than languages, and the gathering of dialects under the wing of a language is another example of political fiat. In practice, once an official language has been proclaimed in a polity, the dialects spoken on the polity's territory and (usually) sharing the official language's dialect continuum are assumed to be the language's dialects. The defining instrument of this process is the state frontier. Hence, although in line with Bloomfield's definition Low German should be a dialect of Netherlandish, it is a dialect of German, as the speakers of Low German live on the territory of Germany.

Kloss described this process of "roofing" a group of dialects by a language through the concept of *Dachsprache* ('roof language'), usually translated as "umbrella language." Every standard or official language is an "umbrella language," as the process of standardization is costly, takes a long time, and is potentially rife with social and political conflicts. Therefore, it is not viable to transform every single dialect into a written language.

In Central and Eastern Europe and on the territory of the former Soviet Union, the process of language-making has been closely connected to the parallel process of nation-making (cf Landau and Kellner-Heinkele 2001). In light of that, the arbitrariness of deciding which human group should be a nation and which dialect a language seems to be of the same ascriptive character (Kamusella 2004). The mechanism of this specifically Central and Eastern European coupling of language and nation is as follows: a nation to be successfully founded and to be considered by other (already recognized) nations and nation-states as legitimate must develop its own specific standard language, considered as the national language. When the nation concerned gains its own polity, or nation-state, this language is usually elevated to the status of the state's sole official language (Kamusella 2001; Kamusella 2006).

4. Evangelical millennialism and languages

One of the legacies of the Reformation is a continuing urge to make the Christian Holy Scripture available in vernaculars, or as we would say it

today, in the languages of people(s) as opposed to the Latin of the narrow layer of the highly educated in post-medieval Western and Central Europe. The origin of such official languages of the European Union as Estonian, Finnish, Latvian, Lithuanian and Slovenian goes back to the 16th-century translations into these languages of the scriptures or extracts from them (Feitsma 1990; Kamusella 2009: 149-326). The renewed wave of translations of the Christian Bible is connected to the West's imperial dominance over the world achieved in the second half of the 19th century. The *mission civilisatrice*, with which the imperial powers justified their colonizing ventures, grew from the earlier religious impulse that had legitimized Europe's maritime empires through (often forced) Christianization. This usually was not officially the case in the 19th century, but in everyday practice the colonial concept of "civilizing natives" was often equated with making "the teachings of Christ" available to them, or, even better, with converting them to Christianity. Christian missions became an indispensible instrument of the colonial *mission civilisatrice* (cf Mont-Rond 1846: x-xi). The vast and multifarious proselytizing effort has been led by American and British evangelical movements (Robert 2003; Stanley 1990). The goal was and remains to accelerate the Second Coming of Christ, an event which, in the theological understanding of these groups, can occur only when every person on earth has been reached with the message of the Gospel (Griswold 2010: 28).

The non-denominational British and Foreign Bible Society, founded in 1804, has been at the forefront of translating the Bible into the world's languages (Why 2011; Zemka 1991). The United States' answer to the British initiative was the American Bible Society, established in 1816, with the identical goal, currently expressed as follows: "The mission of [the] American Bible Society is to make the Bible available to every person in a language and format each can understand and afford, so all people may experience its life-changing message" (Who 2011; Wosh 1994). Similar Bible societies and evangelical organizations with the aim to produce more translations of the Christian Scriptures mushroomed throughout the 19th century, necessitating a coordination of their efforts. To this end, in 1946 they banded together as the United Bible Societies. Since 1950 the UBS has published a specialist journal, *The Bible Translator.* In 2010, the supra-organization's 145 member organizations made the Bible available in 469 languages, and the New Testament in 1,231 languages (Bible 2011).

The Wycliffe Bible Translators, a UBS member, was established in 1942 in the United States, though its global headquarters is now in Singapore. At present it seems to be the most active organization in the field of Bible translation. It maintains that it is the "world's largest Bible translation organization" with 6,400 translators (About Wycliffe 2011; Wycliffe 2011). Its current ambitious objective is as follows, "[b]y the year 2025, together with partners worldwide, we aim to see a Bible translation programme

begun in all the remaining languages that need one" (What 2011). They estimate that 2,565 translations of the Bible, the New Testament, or fragments of them are available for about 4.7 billion people, which means that the Christian Scripture, according to them, must be still translated into 2,078 languages for 340 million people (2010 Scripture). The scope and ambition of the project is such that many accuse it of cultural imperialism, meaning the imposition of values and cultural practices of the rich North on the poor South (Hvalkof and Aaby 1981; Vickers 1984). This imposition frequently translates into religiously-based political, economic and ethnic conflict, often bloody and destructive, but safely (for the rich North) played out in the poor South (Griswold 2010).

5. Accounting for all the world's languages

The Summer Institute of Linguistics (SIL), established in 1934 in the United States, was closely connected to the Wycliffe Bible Translators, first, by their shared founder, the American Christian missionary William Cameron Townsend. Furthermore, SIL's goal is to provide linguistic, anthropological and sociolinguistic expertise to aspiring Bible translators. SIL harnessed linguistics and social sciences for the purpose of improved Bible translation (A Brief 2011). In the early years of SIL the perplexing question was how many languages there were in the world. SIL addressed this problem with its regularly updated publication, *Ethnologue: Languages of the World*. The first edition published in 1951 contained information on a mere 47 languages. The *Ethnologue*'s database was computerized in 1971, and its 16th edition, published in 2009, accounted for 7,358 languages. Since 1974, the hallmark of the *Ethnologue* has been the three letter codes for each language included. The explanation given for this procedure was as follows:

> Each language is given a three-letter code on the order of international airport codes. This aids in equating languages across national boundaries, where the same language may be called by different names, and in distinguishing different languages called by the same name (History 2011).

Recognizing the aforementioned insufficiency of the ISO 639-2 standard, in 2002 the International Organization for Standardization invited SIL to develop an ISO 639-3 standard to cover all the world's languages. The standard was approved in 2007, and SIL was made its registration authority.[11] (Through 2007 it was the ISO 639 Joint Advisory Committee at the Library of Congress that de facto accepted submissions for ISO 639-3 codes.) Incidentally (though perhaps not), to dissociate the Summer Institute of Linguistics somewhat from its confessional origin and mission, its name was changed to the more secular-sounding 'SIL International,' with the "SIL" acronym not spelled out (ISO 693-3 2007; History 2011).

SIL International's approach to allocating codes to languages is quite catholic, and extends to existing and extinct languages, ancient and constructed, those that are endowed with a writing system and those that are exclusively spoken (Request 2011). This code allocation is the actual uniformized world-wide registration of languages that amounts to their de facto international recognition. It also appears to be de jure (though not overtly articulated as such) recognition in light of international law, insofar as the International Organization for Standardization creates and maintains elements and procedures of this law. To my knowledge there are no states or other international relations actors that question the validity of the ISO 639-1, ISO 639-2 and ISO 639-3 Registration Authorities' activities.

Obviously, allocating a code to a language does not entail its recognition in any official function in the state or states where it is (or was) spoken or used in any other way (as in the case of signed languages, or ancient and extinct languages that survive only in the written mode). But this code allocation does have the effect of providing a form of administrative assertion by a duly authorized body on behalf of the global community that a language is a language, not a dialect. As mentioned above, the process of making languages being political in nature and *not* linguistic, the ISO 639 registration of a language follows the clearly articulated will of a speech community to have their dialect recognized as a language. The speech community's elite brings this will to the attention of the world community by applying for an ISO 639 code for their language with one or more of the three Registration Authorities of the standard. At present such applications are in practice limited to requests for ISO 639-2 and ISO 693-3 codes.

6. Evangelicalism and the global regime of language recognition

It seems that the worldwide scope of the ISO 639 standard and the nature of its organization and functioning display all the typical features of a global regime, in this case, a global regime of language recognition. A regime of this kind requires a "global authority" granting recognition of something and an actor that is eligible and able to request such recognition for this "something" (cf Krasner 1983: 2). In this way the regime operates to the preferential advantage of corporate and state actors from the rich North. This is the case because in the poor South, there are very few such actors that are able (conceptually, financially, politically or organizationally) to apply for the recognition of their dialects as languages; furthermore they may favor other public policy objectives, be they popular education or access to clean water for all. However, in the eyes of the rich North (a popular view, espoused by scholars [Statistical Summaries 2009]), the vast majority of the world's identified 7,000-odd languages are spoken in the poor South. A question then arises about what actor or agency effected the registration (and hence, recognition) of

thousands of languages in the South, if the process was not instigated by the relevant speech communities themselves and by their own elites.

It appears that in these cases the initiating agencies have been SIL and other evangelical and Bible societies. The process began in earnest in the 19th century when (mostly evangelical) Christian missionaries ventured deep into colonial and non-Western parts of the world. Their paramount objective being to spread the "good news of Christ," missionaries aspired to translate, first, the New Testament, and later the entire Bible into "local languages." But in most cases, prior to the arrival of missionaries, no local languages existed there as entities in a manner that would conform to the accepted Western paradigm of a language. The vernaculars encountered by missionaries needed to be tweaked into standardized artifacts that would be recognizable as 'languages' to Western scholars, intellectuals and politicians (Errington 2008; Schreyer 2000).

To this end, missionaries acquired the (usually) exclusively spoken dialect of the locality where they happened to have founded their mission. Then they reduced it to writing, predominantly in line with the spelling practices of their native Western languages (mostly Danish, Dutch, English, French, German, Russian[12] or Swedish) or with the practices of scholarship (philology, and later linguistics) as obtaining, at a given time, in their native states. With these instruments in hand, they wrote a grammar and a dictionary of the dialect, making it into "a language" recognizable to the Western eye. In an overwhelming number of cases these grammars and dictionaries did not follow the internal linguistic logic of the non-Western dialect. Western classical models of the two genres, initially developed for Latin in the Roman Empire, and subsequently adopted for Western and Central European vernaculars especially after the Reformation, were followed and taken to be normative, despite any inappropriateness for the linguistic work being undertaken (Hastings 1997: 151-152).

Furthermore, this mission-(pre)ordained making of a dialect into a language was not to meet any actual needs of a concerned speech community, but a missionary organization's self-adopted chiliastic and proselytizing belief in the Christian God-inspired imperative to make the Bible available in all the world's languages. The wholesale destruction of traditional communities and institutions thus executed was of no concern. It went hand in hand with the Western secular administrations' goal of making colonial populaces docile, conforming to the expectations, values, and political and economic needs of the imperial powers. This destruction of communities also entailed the irreversible extinguishing of their languages. Most missionaries praised the often objectively tragic disappearance of the old order, condemning that old order *ex cathedra* as "pagan," and hence not conducive to the Christian vision of "salvation" (cf Achebe 1958; Hochschild 1998).

The process initiated by the missionaries has profoundly complicated the contemporary linguistic landscape in other ways also. If a mission had been established 30 kilometers away from where it actually was built, the language X developed by the missionaries would have been different, often markedly so, from what now has the status of a recognized language. The dialect a missionary took as the basis for his or her pet linguistic project was a result of fortuitous and contingent chance, determined by where it was possible, practical and prudent for the missionary to found his or her mission (Hastings 1997: 152). The choice frequently did not follow any internal socio-political logic obtaining in the region of his or her mission; it is in this respect similar in its linguistic, cultural and political blindness to the arbitrary drawing of the colonial frontiers on the map of Africa at the Berlin Conference of 1884-1885. The missionary linguistic project sometimes sidelined a dominant social group (elite) in a region and pushed into the limelight a group that previously did not display any political ambitions, thus, shattering, fragmenting and reconstituting pre-colonial societies, which facilitated the rule of colonial powers; *divide et impera*, indeed! Soon, however, it also emerged that 'natives' endowed with literacy in their own vernacular by missionaries did not always employ it for purposes envisioned by the latter (cf Olson and Torrance 2001: 123-199). Some also took such linguistic projects away from missionaries' and colonial hands and shaped them to their own liking (cf Peterson 2006).

7. SIL and the methodology of language differentiation

SIL, in deciding what is a language, tacitly follows Bloomfield's impractical distinction between language and dialect that hinges on the concept of mutual intelligibility or lack thereof. In this way two or more dialects, if deemed mutually intelligible, may be conjoined together as a new language, or with an already codified language. Otherwise, an existing language may also be split into two or more, when SIL specialists come to the conclusion that the dialects of some speakers serviced by this language are not mutually comprehensible (or comprehensible enough) with this language (Lewis in Erard 2005). It is substantially a subjective procedure that does not take into consideration the usual extralinguistic (sociopolitical) forces that are commonly at play when it comes to deciding which dialect is a language in the rich North.

> [T]he fewer words from Dialect B that a speaker of Dialect A can understand, the more likely S.I.L. linguists will say that A and B need two Bibles, not one. The entry for the Chadian language of Bernde, for example, rates its similarity to its six neighboring languages from 47 to 73 percent. Above 70 percent, two varieties will typically be called dialects of the same language (Erard 2005).

The elevation to the status of the highest good of the comprehensibility of the produced translations of the Bible to the speakers of all the world's

languages produces paradoxes. The *Ethnologue* records 54 languages in Guatemala, although other sources, including Maya activists, enumerate a mere 18. The Maya movement does not want this mission-produced division of their language of Kaqkchikel into 24 languages (Grinevald in Erard 2005). Another paradox is connected to the very act of research and how it influences the subsequent official perception of the sociolinguistic situation in a given place. For instance, in its coverage of Japan, the Ethnologue breaks Ryukyuan (spoken in Japan's Ryukyuan Islands) into multiple "languages" and "dialects." The former are designated as such on the islands already thoroughly studied by linguists, while the sobriquet of "dialect" tends to be assigned to varieties employed on the islands not yet covered by linguistic surveys.[13]

This Bible translation-based overcounting of languages imposed from outside, is a partial explanation of why there are so many languages in the poor South and so few in the rich North. The same Bible translation methodology of SIL, if it were to be applied to the first-tier of dialects existing in the German-speaking states (Austria, Belgium, Germany, Liechtenstein, Luxembourg and Switzerland), would result in about 40 separate languages[14] (cf König 2005: 138). And if pushed to an extreme it could yield as many as 500 languages (cf Falck 2010). Extending the area under consideration to the entire continent and applying the same criteria, SIL could register as many as 2,000 to 3,000 languages in Europe alone. In 2009, the *Ethnologue* recorded 2,110 languages in Africa spoken by 726 million speakers, while only 234 languages were recorded as being spoken by the 1.55 billion speakers of European languages in Europe and worldwide (as many European languages are spoken widely outside the continent) (Statistical 2009). The huge number of speakers of the European languages, less the speakers of European languages outside Europe, equates to the continent's population, today, 850 million. From the perspective of the methodology deployed by SIL to Africa's languages, and assuming the same distribution of discrete languages, in Europe there should be of the order of 2,300 languages. This number would closely correspond, per the unit of population, to the 2,110 languages in Africa.

SIL's methodology of language differentiation amounts to the proliferation of Ausbau languages within various dialect continua. It does not attempt to alloy Abstand dialects from different dialect continua into one language. Furthermore, SIL refrains from applying its language-differentiation methodology in the rich North or in those parts of the poor South where there have been well established written languages. However, on the basis of SIL's methodology, one could easily argue for the separation of Low German from German, or for splitting the (often Abstand) dialects of Arabic or of Chinese into separate languages in their own right.[15]

Today, many linguists and users at large refer to the *Ethnologue* as the authoritative reference on *all* the world's languages. But it is, as shown above, a quite relative (if not altogether fallible) yardstick. Others praise

SIL for identifying endangered languages and saving them from disappearance. But, at least in some cases, the opposite seems to be the truth:

> It is absurd to think of S.I.L. as an agency of preservation, when they do just the opposite. Note that along with the extermination of native religion, all the ceremonial speech forms, songs, music and art associated with the religion disappear too (Moore in Erard 2005).

The seizure of language research and language-building by Christian missionaries outside in European colonies, and more broadly, outside the West, has taken place largely unnoticed by scholars and public at large. Until recently linguists were rather happy and complacent that missionaries and other enthusiasts of proselytization were doing field research for them. The fact that it was done in the name of this or that strain of Christianity did not perturb scholars overly. But now, having noticed a darker side of this centuries-long Christianizing project, a new field emerged in the scope of which researchers probe into the history and practices of this "missionarly linguistics" (cf SIL International and the disciplinary culture of linguistics 2009).

8. Conclusion: Languages and the brave modern world

For better or for worse, the emerging global regime of language recognition is what it is. One may or may not like its Western ideological and confessional origins and ramifications, but at present it has become an indispensable element of the cyberspace (internet, IT and the like) infrastructure of today's world. The ISO 639-1, ISO 639-2, and especially SIL's ISO 639-3 codes for the world's languages enjoy "near acceptance in library indexing and multilingual software standards" (Erard 2005).

The three standards list all the languages that are of political importance and display a potential for economic exploitation by providing publications, software and information in them. It is so, because from the emic perspective (the insiders' vantage shared by the members of the languages' speech communities), speakers of the languages tend to perceive them as languages in their own right, *not* as dialects or mere language varieties. What is more, they are ready to use the languages in a symbolical manner to define such speech communities as (ethnic) nations or ethnic groups that, by definition, are endowed with certain political rights, and toward which members feel an attachment.

None of the potential competitors poses any imminent threat of dethroning the ISO 639 family of standards. The most serious of these competitors is the Linguasphere Observatory, or the Observatoire, from the originally French name of the organization, *Observatoire Linguistique*. It was founded in 1983 in Canada's Quebec, but four years later moved to France, where it is registered in Normandy. The Observatoire (as it is

popularly referred to in English), hailed as a "transnational language research network," has its academic hubs in Wales and Provence, which places it among other organizations supporting the development of minority languages in the European Union. On the other hand, the late Leopold Sedar Senghor (a Francophone poet, a French deputy, and the first President of Senegal) graced the Observatoire as its honorary president, which has placed this institution in the political-cum-cultural sphere of Paris's cultural-cum-linguistic "empire" of the Francophonie. In this the organization is a secular (though quite culturally and nationally specific) reply to the classification and registration of the world's languages methodologically and institutionally dominated by the Anglophone world and by Bible-centered evangelicals.

The Observatoire's approach to applying codes to language is novel in that it combines genetic and areal classifications, earlier a frequent subject of disagreements among linguists siding with this or that preferred classificatory method. What is more, it goes beyond the political opposition of language vs dialect, and aspires to register "every language variety" in the world. The initial result of this labor of love was the two-volume publication *The Linguasphere Register of the World's Languages and Speech Communities* published in 2000. But the project most probably will not leave its imprint on the emerging global regime of language recognition. First, it went online very late, in May 2011. Second, it registers more than 32,800 codes and 70,900 linguonyms (names of languages), a level of detail that is excessive both for the average user and for the IT industry. Furthermore, the Observatoire enjoys a less secure financial standing than SIL International.

However, the Observatoire's codes have entered Wikipedia and other references, alongside scholarly publications. This organization cooperates with the International Organization for Standardization, and supports and accepts the ISO 639 family of standards. But the Observatoire's system of codes is not a viable alternative to ISO 639. Recently the Observatoire cooperated on the ISO 639-6 standard of four-letter codes that is to record all the globe's "language variants," but GeoLang, not the Observatoire, was made its registration authority in 2009.[16] GeoLang is a Wales-based company that runs this standard on behalf of the British Standards Institution, which is the maintenance agency of ISO 639-6. As a preparation for this monumental task, GeoLang had founded a World Language Documentation Centre in 2007. Tellingly, in contrast to the case of ISO 639-1, ISO 639-2 and ISO 639-3, there is no procedure for requesting an ISO 639-6 code. This suggests that this standard is to remain a preserve of scholars and that the vast majority of extant and recognizable speech communities have no interest in endowing their language varieties with a political meaning (Dalby 2000; GeoLang 2011; ISO 639-6 2009; Welcome 2011). They may not even perceive themselves as separate speech communities in their own right, as they are distinguished as such on the etic basis, that is, by outsiders, scholars

trained to notice and describe language differences that are not endowed with any particular political and societal meaning by speakers themselves. As a result, the overwhelming majority of these varieties are not economically exploitable, because their speakers have no political expectations on their basis, and do not wish to have written information available in them.

How are ISO 639 language codes employed in IT practice? The global cyberspace is run by the Internet Engineering Task Force (IETF), founded in 1986, its mission being "to make the Internet work better." This is an open standards organization, that is, with no formal members, its work done by volunteers (IETF 2011). In 1995 IETF developed language tags (based on ISO 639 codes) for software and internet needs. Six years later (2001), IETF adopted ISO 639-2, and in 2006 ISO 15924 codes for writing systems. In 2009 IETF added ISO 639-3 and ISO 693-5 codes to the repertory of its tags. The Internet Assigned Numbers Authority (IANA), which, in its own words, "is responsible for the operational aspects of coordinating the Internet's unique identifiers and maintaining the trust of the community to provide these services in an unbiased, responsible and effective manner" (IANA 2011), cooperates with IETF on the language tags via the IANA Language Subtag Registry (IANA 2011; Language Tag 2009; Phillips and Davis 2009).

All the aforementioned standardizing, technical and engineering solutions allow for the unencumbered use of over 600 languages in more than 160 scripts on the internet and in a wide variety of software products. With the gradual adoption of the ISO 639-3 standard, there is a possibility for a rapid growth in the number of languages employed by users in cyberspace. Obviously, it depends on users themselves whether they will be interested in reading internet content in their languages and contributing such content to cyberspace. Even in the case of languages with relatively many speakers it is not a foregone conclusion. Multilingualism and polyglossia (employing different languages in different spheres of one's life) are quite widespread outside the rich North; people may speak one language, yet rely quite naturally on information resources published in another language, without detriment to their attachment to or competence in their primary spoken language. For instance, Swahili-speakers tend to use the English Wikipedia, instead of the Kiswahili Wikipedia, which as a consequence remains underdeveloped. In an effort to change this, in 2009 Google sponsored a competition, open to Kenyan and Tanzanian citizens, for the best Swahili-language article in the Swahili Wikipedia, either original or a translation from the English Wikipedia (Kiswahili 2009).

Nowadays, the 281 existing Wikipedias in different languages are the most flexible and extensive multilingual online reference available. With few exceptions the Wikipedias' codes follow those found in the ISO 639-1 and ISO 639-3 lists. Interestingly, due to political differences there are two

Belarusian language Wikipedias, the main difference being the language's two different spelling systems. The Soviet one is associated with the current state administration, and the pre-Soviet (Taraškievica) with the opposition (List 2011a; List 2011b).

The largest Wikipedias boast millions of articles or at least hundreds of thousands; the first group features Wikipedias exclusively in Western languages, while the other mostly in these languages. For instance, the Thai Wikipedia with its 50,000 odd articles has no chance of catching up with the English Wikipedia that is comprised of 3.5 million articles. In January 2011, the international company Asia Online, with its headquarters in Singapore, launched a Thai-language consumer portal. Co-financed and supported by the Thai government the initiative is also a reply to the dilemma of the paucity of Thai-language content on the internet. Among others, the new portal includes machine translations into Thai of all 3.5 million articles from the English Wikipedia (Asia 2011; List 2011b). It still remains to be seen whether statistical machine translation will provide content of a quality acceptable to human readers. Facing such a possibility, some companies have already begun gambling on it, among others, the Open Language Archives Community (OLAC), established in 2000. Its mission is to

> [c]reat[e] a worldwide virtual library of language resources by: (i) developing consensus on best current practice for the digital archiving of language resources, and (ii) developing a network of interoperating repositories and services for housing and accessing such resources (OLAC 2011).

It goes without saying that in its endeavors OLAC sticks to the ISO 639 standards (OLAC Language 2008). Likewise, GeoLang collaborates with the multilingual resource OmegaWiki that aspires to become a dictionary of as many languages as possible. Nowadays it contains "442723 expressions for 46171 concepts in 290 languages" (GeoLang 2011; OmegaWiki 2011). The story continues.

Czissowa and Sapporo
May and August 2011

9. Notes

[1] The article gained much from advice and attention given to it by Michael O Gorman, for which I thank him warmly. I also thank the two anonymous reviewers for their invaluable advice. I bear responsibility for any mistakes and misrepresentations that may remain in the text.
[2] I use the more general term "uniformization," in preference to the usual and technical coinage 'standardization,' in order to emphasize the unprecedented degree of technological, political, cultural and other uniformization across the globe, which is the very product of modernization, however defined.
[3] The United States' NTSC (National Television System Committee), developed in the 1950s, is of no direct interest for the argument here, as its use was and is limited to

Northern America, parts of Latin America, Burma, Japan, the Philippines and Taiwan. In the western hemisphere, insulated from the rest of the world by the US's Monroe Doctrine, there was no communist power that could enter the fray with its own television system to compete with the NTSC on an ideological-cum-technological plane. Indeed, even Cuba adopted the NTSC system. During the Cold War Washington did not oppose Western Europe taking its own technological path, as long as it was not shared with the Soviets. A totalizing uniformization of the technological standards employed in the West was then not yet an issue.

[4] In English and French the name of his organization is abbreviated to ISO, which is not an acronym, but a logo, derived from the Greek word *isos* for "equal." Some disagree with this explanation (Kuert 1997). In this text, for the sake of clarity I stick to the full name of the International Organization for Standardization, because I employ the term "ISO" for numerous standards mentioned throughout.

[5] Obviously, the changes did not necessitate the rise of a Universal Character set in a deterministic manner. It was a choice consciously taken to level out disparate systems, thus allowing for a worldwide uniform cyberspace. What still differs (and is going to continue to do so in the foreseeable future) is whether people have unencumbered access to the net or not.

[6] The RLG, or the Research Libraries Group, is a consortium of US libraries founded in 1974 to computerize and electronically process catalogs and library holdings, and also to work out mutually compatible standards for such procedures and actions. In 1992 the British Library joined this consortium. NeXT (1985-1996) was a US computer company developing and manufacturing computer workstations.

[7] The acronym 'IEC' stands for the International Electrotechnical Commission that since its inception in 1906 has prepared and overseen standards for electrical and eletronic technologies. The commission cooperates closely with the International Organization for Standardization.

[8] The ISO-639-1 list is maintained by Infoterm (International Information Center for Terminology) in Vienna, Austria. The history of Infoterm is connected to the International Organization for Standardization's ISO / TC 37 Technical Committee, founded in 1952. The Committee was entrusted with devising, coordinating and maintaining standards and methods concerning terminology and language resources. Among these was the ISO 639 family of standards. Infoterm was the brainchild of the Austrian self-made industrialist, Esperantist and scholar, Eugen Wüster, who had published his seminal work on technical language standardization in 1931. Because he had already founded his private information center, he was tasked with the administration of the Secretariat of ISO / TC 37 in 1952, and ran it on behalf of the Austrian Standards Institute (ASI) until 1971. In that year Infoterm was founded to continue the work, on the basis of a contract between ASI and Unesco. In 2009 the Secretariat was passed on to China's National Institute for Standardization. Seven years earlier, in 2002, Infoterm had been designated as the Registration Authority of ISO 639-1, but the day-to-day work is conducted by the Joint Advisory Committee (for ISO 639-1 and 639-2), founded in 1998, which is hosted at the Library of Congress (*30 Years of Infoterm* 2004; ISO 639 Joint 2011; ISO Language Codes 2011; ISO/TC 37 2011; Wüster 1931).

[9] To be fair, though, languages and dialects being both of them "emanations" of the human capacity for language, are paradoxical artifacts. On the one hand, they are fashioned by political, social, economic and other cultural forces; on the other they constitute the very cultural medium through which human cultures (or culture in the singular) are created and maintained (Searle 2010: 61-123).

[10] Obviously, it is not to say that culture(s) as created by humans did not leave a mark on dialects. As observed above, dialects change from village to village, from town to town, from region to region, from country to country; and self-evidently it was humans who built or institutionally created all these villages, towns, regions and countries. The physical circumstances altered by human activity and the institutional creations of humans constitute the environment in which human groups live and with which they define themselves and maintain their persisting existence. A degree of permanent

separation thus introduced among human groups, with time, translates itself into language difference, that is, into dialects and dialect continua.

[11] For a comparison of ISO 639, ISO 639-1, ISO 639-2, ISO 639-3, old SIL and other internationally used language codes, see: Language Code (2011).

[12] I include the Russian language in this enumeration, as the Russian Empire, since its founding in 1721, aspired to emulate the Western model of development (or modernization), including the administrative, economic and cultural practices of colonialism (cf Khodarkovsky 2002: 184-219; Miller 2009).

[13] I thank an anonymous reviewer for this useful example.

[14] It was a possibility that Martin Luther clearly wanted to avoid when translating the Bible into German. By consciously shunning terms specific only to this or that dialect, he created a text intelligible to speakers of many dialects. In today's terms we could say that Luther sought to make his German Bible intelligible to the speakers of the large section of the Germanic dialect continuum enclosed within the borders of the Holy Roman Empire. Printers publishing further editions of the Luther Bible contributed to this process by supplying explanations of Luther's less intelligible terms in local dialect counterparts extant in the region for which a given edition was intended (Szulc 1999: 67, 69).

[15] In the terms of Kloss's Abstand-Ausbau opposition, the dialects of Arabic and Chinese that are of an Abstand nature seem to have lost their tendency to go their own separate ways. In these instances two different extralinguistic centripetal forces have ensured cohesion: long-lasting common statehood in the Chinese case and a stable politico-religious commonality in the Arabic one. In both cases these centripetal forces are symbolized by the same written language (and its specific writing system) shared (in the past often merely in a token manner, with no literacy in it on the part of a large segment of a population), which, nevertheless, creates and underpins the pervasive and all-embracing unifying pattern of a respective culture. Hence, these extralinguistic forces, by investing in and being symbolized by the written language, turn linguistically Abstand dialects into culturally and politically Ausbau dialects, susceptible to being "roofed" under "their" umbrella language.

[16] The standard ISO 639-4 (published in 2008) is not a list but a set of guidelines and general rules for language coding. On the other hand, ISO 639-5 (adopted in 2008) develops collective codes for "language families and groups." The list of currently 114 ISO 639-5 codes is made available at the website MultiTree: A Digital Library of Language Relationships (http://multitree.org/), run by the LINGUIST List: International Linguistics Community Online (http://linguistlist.org/).

10. References

2010 scripture access statistics (2011). http://www.wycliffe.net/ScriptureAccessStatistics/tabid/73/language/en-US/Default.aspx (accessed 19 May 2011).

30 Years of Infoterm (2004). Vienna: Infoterm, http://www.infoterm.info/pdf/about_us/30_years_infoterm.pdf (accessed 19 May 2011).

About agency (2011). http://www.gost.ru/wps/portal/pages.en.About?WCM_GLOBAL_CONTEXT=gost/GOST/About+Agency (accessed 16 May 2011).

About Wycliffe Bible Translators USA (2011). http://wyleiffeusa.wordpress.com/about/ (accessed 21 May 2011).

About organization (2011). http://www.easc.org.by/english/mgs_org_en.php (accessed 16 May 2011).

A brief history of SIL International (2011). http://www.sil.org/sil/history.htm (accessed 19 May 2011).

Achebe, Chinua (1958). *Things Fall Apart: The Story of a Strong Man*. London: Heinemann.

Anderson, Chris. 2006. *The Long Tail: Why the Future of Business Is Selling Less of More*. New York: Hyperion

Asia Online Thai Language Portal launch event (2011). http://th.asiaonline.com/launch (accessed 22 May 2011).

Bible translation (2011). http://www.unitedbiblesocieties.org/?page_id=207 (accessed 19 May 2011).

Blackburn, Bonnie and Holford-Strevens, Leofranc (1999). *The Oxford Companion to the Year: An Exploration of Calendar Customs and Time-Reckoning*. Oxford: Oxford University Press.

Bloomfield, Leonard (1926). A set of postulates for the science of language. *Language* 1(2): 153-164.

Ciscel, M H (2006). A separate Moldovan language? The sociolinguistics of Moldova's Limba de Stat. *Nationalities Papers* 34(5): 575-597.

Codes for the representation of names of languages (2011). http://www.loc.gov/standards/iso639-2/php/English_list.php (accessed 16 May 2011).

Codes for the Representation of Names of Scripts / Codes pour la représentation des noms d'écritures (2011). http://unicode.org/iso15924/codelists.html (accessed 16 May 2011).

Craig, Murphy (1994). *International Organization and Industrial Change: Global Governance since 1850*. Oxford: Oxford University Press.

Crane, Rhonda J (1979). *The Politics of International Standards: France and the Color TV War*. Norwood NJ: Ablex Publishing Corporation.

Dalby, David (2000). *The Linguasphere Register of the World's Languages and Speech Communities* (2 vols). Hebron, Wales: published for Observatoire Linguistique by Linguasphere Press/Gwasg y Byd Iaith.

Durand, Olivier (2009). *Dialettologia araba*. Rome: Carocci.

Erard, Michael (2005). How linguists and missionaries share a Bible of 6,912 Languages. *New York Times*, 19 July, http://www.nytimes.com/2005/07/19/science/19lang.html (accessed 18 May 2011).

Errington, J Joseph (2008). *Linguistics in a Colonial World: A Story of Language, Meaning, and Power*. Malden MA and Oxford: Blackwell.

Falck, Oliver, Stephan Heblich. Alfred Lameli. Jens Südekum (2010). *Dialects, Cultural Identity, and Economic Exchange* (IZA Discussion Paper 4743). Bonn: Forschungsinstitut zur Zukunft der Arbeit.

Feitsma, Anthonia (1990). The Reformation and the vernacular. In *Fourth International Conference on Minority Languages*, (Vol 2: Western and Eastern European Papers), Durk Gorter, Jarich F Hoekstra, Lammert G Jansma, and Jehannes Ytsma (eds.), 1-10. Clevendon, Avon: Multilingual Matters.

Ferguson, Charles Albert (1959). Diglossia. *Word*. No 15(1): 325-340.

GeoLang (2011). http://www.geolang.com/index.php (accessed 22 May 2011).

Goldman, D T and R J Bell (eds.) (1986). *The International System of Units (SI)*, (National Bureau of Standards, Special Publication 330). Washington: US Government Printing Office.

Gooskens, Charlotte (2006). Mutual comprehensibility of written Afrikaans and Dutch: Symmetrical or asymmetrical?. *Literary and Linguistic Computing* 21(4): 543-557.

Greenberg, Robert D (2004). *Language and Identity in the Balkans: Serbo-Croatian and Its Disintegration*. Oxford: Oxford University Press

Griswold, Eliza (2010). *The Tenth Parallel: Dispatches from the Fault Line Between Christianity and Islam*. New York: Farrar, Straus and Giroux.

Hastings, Adrian (1997). *The Construction of Nationhood: Ethnicity, Religion and Nationalism*. Cambridge: Cambridge University Press.

Haugen, Einar (1966). Semicommunication: The language gap in Scandinavia. *Sociological Inquiry* 36(2): 280-297.

History of the *Ethnologue* (2011). http://www.ethnologue.com/ethno_docs/introduction.asp#history (accessed 19 May 2011).

Hochschild, Adam (1998). *King Leopold's Ghost: A Story of Greed, Terror, and Heroism in Colonial Africa*. Boston: Houghton Mifflin.

Hvalkof, Soren and Peter Aaby (eds.) (1981). *Is God an American? An Anthropological Perspective on the Missionary Work of the Summer Institute of Linguistics*. Copenhagen: IWGIA and Survival International.

IANA (2011). http://www.iana.org/ (accessed 22 May 2011).

ICANN approves internet addresses in Arabic, Cyrillic Script (2010). *The Independent*, Jan 24. http://www.independent.co.uk/news/media/icann-approves-internet-addresses-in-arabic-cyrillic-script-1877447.html (accessed 16 May 2011).

IETF (2011). http://www.ietf.org/ (accessed 22 May 2011).

Information I33 -1997: Changes to the terrestrial television systems in Central and East European Countries (1997). http://tech.ebu.ch/docs/i/i033.pdf (accessed 16 May 2011).

International Conference. Message from the President of the United States Transmitting a Communication from the Secretary of State, Relative to International Conference Held at Washington (International Conference Held at Washington for the Purpose of Fixing a Prime Meridian and a Universal Day. October, 1884) (U.S. House Documents. 48th Congress. 2d Session. no. 14.) (1884). Washington: Gibson Bros.

Intner, Sheila S (2007). The passing of an era. *Technicalities* 27(2): 13.

ISO. 2011. http://www.iso.org/iso/home.htm, May 16.

ISO 15924 Registration Authority (2011). http://unicode.org/iso15924/ (accessed 17 May 2011).

ISO 639-3: 2007 (2007). http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=39534 (accessed 19 May 2011).

ISO 639-6: 2009 (2009). http://www.iso.org/iso/catalogue_detail?csnumber=43380 (accessed 22 May 2011).

ISO 639 Joint Advisory Committee (2011). http://www.loc.gov/standards/iso639-2/iso639jac.html (accessed 19 May 2011).

ISO 639 Joint Advisory Committee: Working principles for ISO 639 maintenance (2000). http://www.loc.gov/standards/iso639-2/iso639jac_n3r.html (accessed 16 May 2011).

ISO language codes (2011). http://www.infoterm.info/standardization/iso_639_1_2002.php (accessed 16 May 2011).

ISO/R 639: 1967 (2011). http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=4765 (accessed 16 May 2011).

ISO/TC 37 (2011). http://www.infoterm.info/standardization/iso_tc_37.php (accessed 19 May 2011).

Jensen, John B (1989). On the mutual intelligibility of Spanish and Portuguese. *Hispania* 72(4): 848-852.

Kamusella, Tomasz (2001). Language as an instrument of nationalism in Central Europe. *Nations and Nationalism* No 7(2): 235-252.

Kamusella, Tomasz (2004) On the similarity between the concepts of nation and language. *Canadian Review of Studies in Nationalism* 31: 107-112.

Kamusella, Tomasz (2006). The isomorphism of language, nation, and state: The case of Central Europe (pp 57-92). In *Nationalisms Across the Globe: An Overview of Nationalisms of State-Endowed and Stateless Nations* (Vol 2: The World), Wojciech Burszta, Tomasz Kamusella, and Sebastian Wojciechowski (eds.), 57-92. Poznań, Poland: Wyższa Szkoła Nauk Humanistycznych i Dziennikarstwa.

Kamusella, Tomasz (2009). *The Politics of Language and Nationalism in Modern Central Europe*. Basingstoke: Palgrave.

Kashubian keyboard layout (2011). http://en.kaszubia.com/computers-internet/kashubian-keyboard-layout/windows/kashubian-keyboard-layout-for-ms-windows-2000-xp-vista (accessed 19 May 2011).

Khodarkovsky, Michael (2002). *Russia's Steppe Frontier: The Making of a Colonial Empire, 1500-1800*. Bloomington and Indianapolis: Indiana University Press.

Kincaid, Peter (1986). *The Rule of the Road: An International Guide to History and Practice*. New York: Greenwood Press.

Kiswahili Wikipedia Challenge (2009). http://www.google.com/events/kiswahili-wiki/terms.html (accessed 22 May 2011).

Klassifikator gosudarstvennykh standartov SSSR [The Classifying Handbook of the State Standards of the USSR] (1978). Moscow: Izdatel'stvo standartov.

Kloss, Heinz (1967). Abstand languages and Ausbau languages. *Anthropological Linguistics* 9(7): 29-41.

König, Werner (2005). *dtv-Atlas Deutsche Sprache*. Munich: dtv.

Krasner, Stephen D (1983). Structural Causes and Regime Consequences: Regimes as Intervening Variables (pp 1-22). In: Stephen D Krasner, ed. *International Regimes*. Ithaca NY: Cornell University Press.

Kuckenburg, Martin (2006). *Pierwsze słowo* [The First Word], (Seria Ceramowska). Warsaw: PIW.

Kuert, Willy (1997). The founding of ISO. http://www.iso.org/iso/founding.pdf (accessed 19 May 2011).

Künstler, Mieczysław Jerzy (2000). *Języki chińskie* [The Chinese Languages]. Warsaw: Dialog.

Laks, Bernard (ed) (2008). *Origin and Evolution of Languages: Approaches, Models, Paradigms*. London and Oakvile CT: Equinox.

Landau, Jacob M and Barbara Kellner-Heinkele (2001). *Politics of Languages in the Ex-Soviet Muslim States: Azerbayjan, Uzbekistan, Kazakhstan, Kyrgyzstan, Turkmenistan and Tajikistan*. London: Hurst.

Language code (2011). In *Wikipedia*. http://en.wikipedia.org/wiki/Language_code (accessed 21 May 2011).

Languages and scripts (2011). http://unicode.org/repos/cldr-tmp/trunk/diff/supplemental/languages_and_scripts.html (accessed 16 May 2011).

Language Tag Registry (ITRU) (2009). http://www.ietf.org/wg/concluded/ltru.html (accessed 22 May 2011).

Latimer, Jack (1997). *Friendship Among Equals: Recollections from ISO's First Fifty Years*. Geneva: ISO Central Secretariat.

List of Wikipedias (2011a). http://meta.wikimedia.org/wiki/List_of_Wikipedias (accessed 22 May 2011).

List of Wikipedias (2011b). http://en.wikipedia.org/wiki/Wikipedias (accessed 22 May 2011).

MARC code list for languages (2007). Washington DC: Library of Congress. http://www.loc.gov/marc/languages/langhome.html (accessed 16 May 2011).

Miller, Alexey (2009). "Russifications"? In search for adequate analytical categories. In *Imperienvergleich. Beispiele und Ansätze aus osteuropäischer Perspektive. Festschrift für Andreas Kappeler*, Guido Hausmann, (ed.), 123-144. Wiesbaden: Harrassowitz.

Mont-Rond, Maxime de (1846). *Missions du Levant, d'Azie et de la Chine*. Lille: L Lefortt, Imprimeur-Libraire.

Nábělková, Mira (2007). Closely-related languages in contact: Czech, Slovak, "Czechoslovak". *International Journal of the Sociology of Language*. 183: 53-73.

OLAC (2011). http://www.language-archives.org/ (accessed 22 May 2011).

OLAC language extension (2008). http://www.language-archives.org/REC/language.html (accessed 22 May 2011).

Olson, David R and Nancy Torrance (eds.) (2001). *The Making of Literate Societies*. Malden MA and Oxford UK: Blackwell.

OmegaWiki (2011). http://www.omegawiki.org/ (accessed 25 May 2011).

Peterson, Derek R (2006). Language work and colonial politics in Eastern Africa: The making of Standard Swahili and 'School Kikuyu'. In *The Study of Language and the Politics of Community in Global Context,* David L Hoyt and Karen Oslund (eds.), 185-214. Lanham MD: Lexington Books (a division of Rowman & Littlefield Publishers).

Price, Derek John de Solla (1963). *Little Science, Big Science*. New York: Columbia University Press.

Phillips, A and Davis, M (eds.) (2009). Tags for identifying languages. http://www.ietf.org/rfc/rfc5646.txt (accessed 22 May 2011).

Request for new language code element in ISO 639-3 (2011). http://www.sil.org/iso639-3/ISO639-3_NewCodeRequestForm.doc (accessed 20 May 2011).

Robert, Dana L (2003). *Occupy Until I Come: A T Pierson and the Evangelization of the World*. Grand Rapids MI: Eerdmans Publishing.

Schreyer, Rüdiger (2000). "Savage" languages in eighteenth-century theoretical history of language. In *The Language Encounter in the Americas, 1492-1800: A Collection of Essays,* (European Expansion and Global Interaction), Edward G Gray and Norman Fiering, (eds.), 310-326. New York and Oxford: Berghahn.

Searle, John R (2010). *Making the Social World: The Structure of Human Civilization*. New York: Oxford University Press.

SIL International and the Disciplinary Culture of Linguistics (2009). *Language*. No 85(3): 618-658.

Stanley, Brian. 1990. *The Bible and the Flag: Protestant Missions and British Imperialism in the Nineteenth and Twentieth Centuries*. Leicester: Apollos.

Statistical summaries (2009). http://www.ethnologue.com/ethno_docs/distribution.asp?by=area (accessed 21 May 2011).

Summary narrative (2011). http://www.unicode.org/history/summary.html (accessed 16 May 2011).

Szulc, Aleksander (1999). *Odmiany narodowe języka niemieckiego. Geneza – rozwój – perspektywy* [The National Varieties of the German Language: The Origins, Development and Perspectives], (Rozprawy Wydziału Filologicznego 70). Cracow: Polska Akademia Umiejętności.

The Unicode standard: A technical introduction (2011). http://www.unicode.org/standard/principles.html#, (accessed 16 May 2011).

UTF-8 history (2011). http://www.cl.cam.ac.uk/~mgk25/ucs/utf-8-history.txt (accessed 16 May 2011).

Vickers, William (1984). Review: Fishers of Men or Founders of an Empire? The Wycliffe Bible Translators in Latin America by David Stoll. *American Ethnologist*. No 86(1): 200-201.

Weinreich, Max (1945). Der YIVO un di problemen fun undzer tsayt [The YIVO and the Problems of Our Times]. *YIVO Bleter*. No 25(1): 3-18.

Welcome to Linguasphere.info (2011). http://www.linguasphere.info/ (accessed 22 May 2011).

What is Wycliffe's goal? (2011). http://wycliffe.org.uk/wycliffe/about/faq.html (accessed 19 May 2011).

Who we are (2011). http://www.americanbible.org/about/who-we-are (accessed 19 May 2011).

Why we exist (2011). http://www.biblesociety.org.uk/about-bible-society/why-we-exist/ (accessed 19 May 2011).

Wicherkiewicz, Tomasz (2003). *The Making of a Language: The Case of the Idiom of Wilamowice*. Berlin and New York: Mouton de Gruyter.

Wodak, Ruth, Barbara Johnstone, and Paul E Kerswill (eds.) (2011). *The SAGE Handbook of Sociolinguistics*. London: SAGE.

Wosh, Peter (1994). *Spreading the Word*. Ithaca NJ: Cornell University Press.

Wüster, Eugen (1931). *Internationale Sprachnormung in der Technik, besonders in der Elektrotechnik. Berlin: VDI-Verlag*.

Wycliffe Global Alliance: Who we are (2011). http://www.wycliffe.net/AboutUs/tabid/57/language/en-US/Default.aspx (accessed 21 May 2011).

Wyniki Narodowego Spisu Powszechnego Ludności i Mieszkań 2002 w zakresie deklarowanej narodowości oraz języka używanego w domu (2008). Warsaw: GUS. http://www.stat.gov.pl/gus/5840_4520_PLK_HTML.htm (accessed 19 May 2011).

Zemka, Sue (1990). The Holy Books of Empire: Translations of the British and Foreign Bible Society (pp 102-137). In *Macropolitics of Nineteenth-Century Literature: Nationalism, Exoticism, Imperialism*, Jonathan Arac and Harriet Ritvo (eds.), 102-137. Philadelphia: University of Pennsylvania Press.

Zammit, Martin (2000). Arabic and Maltese cognate roots. In *Proceedings of the Third International Conference of Aida*, *Association Internationale de Dialectologie Arabe Held in Malta 29 March-2 April 1998*, Manwel Mifsud (ed.), 241-245. Malta: Salesian Press.