MINISTRY OF THE HIGHER AND SECONDARY SPECIAL EDUCATION OF THE REPUBLIC OF UZBEKISTAN

SAMARKAND STATE INSTITUTE OF FOREIGN LANGUAGES

Report

Words of English through the Internet and other technologies

Done by: Mahmudova Y. 306-guruh

Checked by: Yakubov J.

Samarkand – 2016

PLAN:

INTRODUCTIO

What is the Internet?

English is the main language of communication worldwide

Words of English through the Internet and other technologies

The advantages of using English through

CONCLUSION

BIBLIOGRAPHY

Introduction

On December 10, 2012 President of the Republic of Uzbekistan Islam Karimov signed a decree “On measures to further improve foreign language learning system”.

It is noted that in the framework of the Law of the Republic of Uzbekistan "On education" and the National Programme for Training in the country, a comprehensive foreign languages’ teaching system, aimed at creating harmoniously developed, highly educated, modern-thinking young generation, further integration of the country to the world community, has been created. During the years of independence, over 51.7 thousand teachers of foreign languages graduated from universities, English, German and French multiмedia tutorials and textbooks for 5-9 grades of secondary schools, electronic resources for learning English in primary schools were created, more than 5000 secondary schools, professional colleges and academic lyceums were equipped with language laboratories.[1;қарор]

Generally speaking, English is the universal language on the Internet, but it has no official status, and it will never have. The reasons for the position of English are the imperialism and economical and political importance of English-speaking countries. Linguistically, English is extremely unsuitable for international communication, and the actual wide use of English tends to polarize the world into Internet users and Internet illiterates.

The position of English can only be altered by major world-scale political and economical changes, such as increasing importance of the European Union or a coalition between Japan and China. Such powers might wish and be able to promote a language other than English, possibly a constructed language, for international communication.

What is Internet ?

The dominance of English in the Internet needs no arguing for. Computers are in any case English-oriented. Netscape and Java are in English, the vocabulary of computing and of the Internet is overwhelmingly English, and most of the texts that are accessed through it are in English. The search engines are in English and are in the US. The reasons for the dominance of English are firstly historical -- the Internet began in the USA, which is still the leading user of it, and the USA is an English-using nation. The norms of the Internet are established in ASCII(Anglophone Script) texts, and even now texts transmitted unaltered from (for example) Francophone keyboards may produce garbage on English-favouring keyboards. To avoid this from happening, those who post, for example, French on ASCII lists, must omit diacritics.

Ces troisiemes rencontres, organisees par les etudiants de l'Atelier des Doctorants de Linguistique, se veulent un lieu de contacts et d'echanges fructueux entre doctorants et jeunes chercheurs relevant de differents cadres theoriques et de differents domaines de recherche en linguistique. Chaque communication sera suivie de 10 min. de discussion. D'autre part, des ateliers-debats seront organises le samedi 7 au matin.

There are many accents missing here. It should begin, "Ces troisièmes rencontres, organisées par les étudiants...".

Two things constrain the use of languages other than English on world-wide web documents and communications:

the difficulty of writing languages using non-ASCII characters, and characters that have diacritics

the desire to reach as large a readership as possible.

The first of these constraints can be overcome, though not necessarily with ease. The second constraint, however, means that most documents in languages other than English, including those in countries where English is little used internally are mirrored by English translations.

The world wide web is in the position of a permanent international conference, where papers are either in English or are accompanied by English translations. The role of English reflects the dominance of English in cross-national communication.

Access to the Internet is empowering. It gives access to uncensored material of enormous variety. Access to a world-size (though rather messy) encyclopedia. Access to almost instant communication with individuals located thousands of miles away. A person using the Internet can adopt a new persona -- you can change gender, acquire beauty, lose disability, develop aggression..... There is information available, entertainment, insight into other cultures. Various governments (including those of the USA, France, and Singapore) have attempted to control some of this information, either in terms of what can be placed in URLs (Univesal Resource Locators) or in terms of what can downloaded. But so far no effective way of controlling the reading of documents placed on the web has been found. We may be in a Golden Age of Net use -- when controls become more effective and when more sites are either commerically motivated or no longer cost-free, we will lose the present sense of uncernsored liberty.

However, this empowerment has on the whole been extended most to those who are already most empowered. To gain good access to the web three things are needed:

to live in a location that supports the technology needed to gain access

to have access to a service provider

to be able to use English

There was a time in the early 90s when many countries did not possess servers that were linked to the web. Now, this access at a national level is so widespread that the progress of Balkan politics can be followed on the web. Some countries are still not listed, however. De Mello's College and University Home Pages lists universities in the following 87 countries:

Argentina, Armenia, Australia, Austria, Bahrain, Bangladesh, Belgium, Bermuda, Brazil, Brunei Darussalam, Bulgaria, Canada, Chile, China, Colombia, Costa Rica, Croatia, Cyprus, Czech Republic, Denmark, Ecuador, Egypt, Estonia, Ethiopia, Faroe Islands, Fiji, Finland, France, Germany, Greece, Guam, Hong Kong, Hungary, Iceland, India, Indonesia, Ireland, Israel, Italy, Jamaica, Japan, Jordan, Korea, Kuwait, Latvia, Lebanon, Lithuania, Luxembourg, Macau, Macedonia, Malaysia, Malta, Mexico, Morocco, Netherlands, New Zealand, Nicaragua, Norway, Palestine, Panama, Peru, Philippines, Poland, Portugal, Puerto Rico, Romania, Russia, Singapore, Slovakia, Slovenia, South Africa, Spain, Sri Lanka, Sudan, Sweden, Switzerland, Taiwan, Thailand, Turkey, Ukraine, United Arab Emirates, United Kingdom, United States, Uruguay, Venezuela, Yugoslavia, Zambia

Given that universities have been leaders in Internet access, absence from this list suggests a country either has no university, or has little or no Internet access. Even many countries on this list (e.g. China) have relatively limited access. The poor representation of Africa is noticeable, with only Egypt, Morocco, South Africa and Zambia present. Europe, both East and West, is well covered (Albania is missing, however), as is both North and South America. In Asia, Burma, Laos, Cambodia and Vietnam do not appear. In other words, the poorest countries of the world still are deprived of Internet access.

In many countries which do have Internet access, large portions of the population live in rural areas away from the electricity and telephone provision that supports this basic access, or are not wealthy enough to have such facilities themselves. Even within wealthy countries with effective electricity and telephone systems, the nearest service provider may be sufficiently far from some rural users to make accessing the web prohibitively expensive, due to the cost of long distance or satellite telephone calls.

Private access to the Internet has developed over the 1990s, although there are still technical and commercial problems in private accounts, as the recent blocakages at America Online have shown. Many countries have been slow to provide private access to the Internet. Even where private access is possible, it is costly, with rental fees and telephone costs being added to very substantial hardware costs. In the UK, for example, the average cost of a computer is around £2000. The average weekly earnings of full-time employees on adult rates is £336.30. Thus someone in full time work would have to spend 6 weeks' salary on a computer. The lower the national and individual salary, the more this cost becomes out of reach. In countries with lower average salaries, computers tend to be more expensive both in real and proportional terms.

The first two things needed for Internet access make it an advantage:

to be in a country with good electricity supply, good telephonic connections and the political will to allow Internet access

within those countries, to be an academic user, a user with other corporate access, or wealthy enough to afford a private account

These requirements, essentially economic in foundation, point to countries with high GNP, and to an urban environment within them.

However, the third thing helpful for Internet access, the ability to use English, divides the world into rather different groupings. Sociolinguists (e.g. Moag) have traditionally grouped countries into three categories according to the dominant pattern of English use within them:

English as a native language: Most people grow up speaking English at home. English the dominant language of daily usage for most people. (examples: UK, USA, Australia) (ENL)

English as a second language: Those who know English mostly learnt it after they had learned another language. English often learnt at school. English is used (by

varying proportions of the population) in everyday and in official usage, with citizens often speaking to each other in English.(examples: India, Singapore, Ghana, Philippines) (ESL)

English as a foreign language: Those who know English mostly learnt it at school. English is hardly ever spoken within the country (outside English lessons). (Examples: France, Japan, Thailand) (EFL)

English is the main language of communication worldwide .

In general, the universal language on the Internet is English, or more exactly a vague collection of languages called "English" because their common origin is the national language spoken in England by the English. That national language has spread over the world, and several variants such as American (US) English, Australian English, etc exist. A great number of people whose native language is none of the variants know English as a foreign language. They typically use a more or less simplified variant, e.g. excluding most of the idioms of British, American, Australian etc English. Of course, they make mistakes, and sometimes the "English" used by people as a foreign language on the Internet is almost incomprehensible to anyone else. In addition, people who use English as their native language do not know how to spell difficult words, since they basically know English as a spoken language.

Thus, roughly speaking, the universal language of the Internet is clumsy, coarse and misspelled "English".

There are exceptions, most importantly national newsgroups in such countries where English is not the native language of the majority. Even in such groups English is used, for instance when people from other countries wish to participate. And a few international groups have a theme like discussing a particular language or culture so that it is natural to assume that the participants have a common language other than

English. Moreover, a group which is partly international in the sense of not being purely national might use a language other than English, for instance if the group is intended for people in German-speaking countries.

Why is it so?

Generally speaking, when a languages has got the position of a universal language, the position tends to be affirmed and extended by itself. Since "everyone" knows and uses English, people are almost forced to learn English and use it, and learn it better.

Even if you expect the majority of your readers to understand your native language, you may be tempted to use English when writing e.g. about research work. Usually researchers all over the world know English and use it a lot, and often the relevant terminology is more stable and well-known in English than in your own language. Thus, to maximize the number of interested people that can understand your text, you often select English even if the great majority of your readers have the same native language as you. Alternatively, you might write your texts both in your native language and in English, but this doubles the work needed for writing your document and possibly maintaining it. The maintenance problem is especially important for documents on the World Wide Web - the information system where one crucial feature is the ability to keep things really up to date. Consequently, the use of English in essentially national contexts tends to grow.

In the news system, the position of English in most international groups is regarded as so obvious that people who post non-English articles to such groups - by accident or by ignorance - typically get flamed quickly. This is the sort of control that newsgroup communities exercise in other matters than language, too. It is often regarded as an example of the "democratic" nature of the news system. However, things are changing fast. The flame control - i.e. control by flaming - has worked to a great extent because people have had relatively compatible status, background, and values of life. The time has come - childhood's end for the Internet - when we will see

more and more people and organizations who pay little attention to flames and netiquette. The increasing amount of spamming is just one indication of this. But these general remarks hardly apply to the status of English.

By the way, when people post articles to international groups in their own languages, the reason is typically novice users' ignorance of basic facts about the news system. People start posting articles before they have read what is generally written to the group. One thing that causes this happen relatively often that there is no easily accessible and useable list of groups together with their content descriptions, and typically content descriptions do not explicitely state what language(s) should be used in the group.

The universal language position, once gained, tends to be strong. But how is such a position gained?

During the history of mankind, there have been several more or less universal languages or *lingua francas*, such as Latin (and Greek) in the Roman empire, mediaeval Latin in Western Europe, later French and English. Universality is of course relative; it means universality in the "known world" or "civilized world", or just in a large empire. No language has been really universal (global), but the current position of English comes closest. The position of a universal language has always been gained as a by-product of some sort of imperialism: a nation has conquered a large area and more or less assimilated it into its own culture, including language, thus forming an empire. Usually the language of the conquerer has become the language of the state and the upper class first, then possibly spread over the society, sometimes almost wiping out the original languages of the conquered areas. Sometimes - especially in the Middle Ages - the imperialism has had a definite cultural and religious nature which may have been more important than brute military and economic force.

As regards to the English language, it would have remained as a national language of the English, had it not happened so that the English first conquered the rest of the British Isles, then many other parts of the world. Later, some English colonies in a relatively small part of America rebelled, formed the United States of America, and expanded a lot. They formed a federal state where a variant of the English language was one of the few really uniting factors. And that federal state became, as we all know, wealthy and important. It also exercised traditional imperialism, but more importantly it gained a very important role in world economy and politics. Whether you call the US influence imperialism or neo-imperialism is a matter of opinion, but it certainly has similar effects on maintaining and expanding the use of English as classical imperialism.

This probably sounds like political criticism, but it is intended to be descriptive only. Personally, I do not regard imperialism as an incarnation of the Evil; it has had both positive and negative effects, and in many cases imperialism has been a necessary step from chaos to civilization.

The importance of the Internet grows rapidly in all fields of human life, including not only research and education but also marketing and trade as well as entertainment and hobbies. This implies that it becomes more and more important to know how to use Internet services and, as a part of this, to read and write English.

Of course, the majority of mankind cannot use the Internet nowadays or in the near future, since they live in countries which lack the necessary economical and technological infrastructure. But the Internet causes polarization in developed countries, too: people are divided into Internet users and Internet illiterates, and as the use of the Internet grows and often replaces traditional methods of communication, the illiterates may find themselves in an awkward position.

In general, it is easy to learn to use Internet services. The worst problems of Internet illiteracy are, in addition to lack of economical resources of course, wrong

attitudes. Older people are usually not accustomed to live in a world of continuous and rapid change, and they may not realize the importance of the Internet or the easiness of learning to use it.

But although Internet services themselves are, generally speaking, easy to learn and use, you will find yourself isolated on the Internet if you are not familiar with English. This means that knowledge or lack of knowledge of English is one of the most severe factors that cause polarization. Learning to use a new Internet service or user interface may take a few hours, a few days, or even weeks, but it takes years to learn a language so that you can use it in a fluent and self-confident manner. Of course, when you know some English, you can learn more just by using it on the Internet, but at least currently the general tendency among Internet users is to discourage people in their problems with the English language. Incorrect English causes a few flames much more probably than encouragement and friendly advice.

In different countries and cultures, English has different positions. There are countries where English is the native language of the majority, there are countries where English is a widely known second language, and there are countries where English has no special position. These differences add to the above-mentioned polarization. Specifically, it is difficult for people in previous colonies of other countries than Great Britain (e.g. France, Spain, the Netherlands) to adapt to the necessity of learning English. Locally, it may be necessary to learn the language of the previous colonial power since it is often an official language and the common language of educated people; globally, English is necessary for living on the Internet. And the more languages you have to learn well, the less time and energy you will have for learning other things.

There is no conceivable way in which any authority could define an official language for the Internet. The Internet as a whole is not controlled by anyone or

anything, and this could only change if, by miracle, all countries made an agreement on it or if the entire world were taken to the control of one government.

Thus, if the question "whether or not English should be made the universal language of the internet" is interpreted as concerning the **official** status of English, the answer is simply that English, or any other language, **cannot** be made the official universal language. It is fruitless to ask whether an impossible thing should be made. Things can change, and they actually do, often with unpredictable speed. The rapid fall of the Soviet empire - including the loss of the role of Russian as a "universal" language within in - is a recent indication of this.

English can lose its position as a widely used (although not official) universal language in two ways. Either a new empire emerges and its language becomes universal, or a constructed language becomes very popular. I believe most people regard both of these alternatives as extremely improbable, if not impossible. Perhaps they are right, perhaps not.

I can see two possible empires to emerge: the European Union and a yet nonexistent Japanese-Chinese empire.

The European Union (EU) is an existing formation which is, at least according to its own doctrine, moving towards federalism. In many respects, the European Union already **is** a federal state, with less independence and autonomy for its constituents than the states have in the United States. Although people may present the EU as the successor of previous empires such as the Roman empire and the empire of Charlemagne, it is quite possible that the EU never becomes a real empire, since it seems to be inherently bureaucratic. Every empire needs a bureaucracy, of course, to promote the aims of its ruler(s), but the EU lacks true rulers. But if the EU ever becomes a true empire with prominent role in the world, the language of the empire will hardly be any of the national languages in the EU, except possibly English. It is more probable that the builders of the empire will realize the need for a

relatively neutral universal language, and adopt Esperanto or some other constructed language for official purposes. In fact, such a choice would be extremely rational at the present stage of the EU, since now a considerable portion of EU expenses are used for translation and interpretation between the official languages of the EU. A single official language of the EU might or might not be adopted by people worldwide as a universal language for everyday communication, including communication on the Internet.

Traditional secretarial contractions and acronyms such as **wrt** (with respect to) now share the stage with freshly minted expressions and a host of trite acronyms such as **<g>** (for grin), **LOL** (laughing out loud), or **TTYL** (type to you later) in an attempt to add human warmth and personality to otherwise utilitarian text. This attempt is laudable, even understandable; however, such humanity and emotion can be added to writing by the use of *words*. The use of these expressions, coupled with the penchant of users to omit most standard conventions of the English language, namely punctuation, spelling, syntax, and capitalization, often makes what they write difficult, if not impossible, to comprehend.

The state of writing in Internet communications may also be due to what can be called the "invisible person" syndrome. Accountability breeds meticulousness, while anonymity breeds indolence. Because poorer grade writing skills are the accepted "standard" of the Net communications, people may consider it an excuse, or an incentive, to be less conscientious about espied petty issues such as grammar and spelling. In addition, the perceived anonymity of users on the Net begs the question *why strive for perfection in the English language when no one knows or cares who you are?* Conversely, *who wants to be singled out for doing shoddy work?* These people may be surprised to learn that most users can be traced back to their Internet Service Providers (ISPs) while online by anyone who takes the time to do so. Because Internet communications are the only source from which the online community can

derive an impression of any user, that impression is directly related to the language, tone, and structure employed by the person in question.

A lack of training in the use of, and knowledge about, Internet tools may also contribute to the nature of Internet communications. Consider that the naval explorers of old who struck out into the unknown reaches of the Atlantic in order to discover new lands did not have maps of the regions into which they were sailing, but instead had experience in naval navigation, knowledge of traversing rough seas, and seasoned, experienced crews at their disposal. The modern explorer of the Internet is set adrift upon a vast sea of information without the knowledge of, or experience with, the tools at his or her disposal. Such people may (understandably) seek to cut corners in areas in which they have some mastery (English language), so that more time can be spent learning about areas in which they are less skilled (the Internet).

As new technologies become part of everyday life, it only makes sense that the language change as well. The advent of the Industrial Revolution brought many new words into the English vernacular. The turn of the century brought many more. Just as a medieval chariot became *carriage*, then *horseless carriage*, and finally *car*, so too have words and concepts that were formerly the exclusive bailiwick of computer scientists and other technical professionals entered the layperson's dictionary of common words. This absorption of technical words into the language has resulted in some meanings becoming skewed. In everyday conversations unrelated to computer technology you might hear people talking about "being online" or needing "down time," things "crashing" and machinery "booting." The more terms are incorrectly used, the more the skewed meanings of the words gain legitimacy.

In the end, the most common usage of words and expressions win out, and the English language and the way we use it will be altered to incorporate them. This is how language evolves. Hopefully, the new words and meaning that are constructed

will adequately describe the concepts necessary for effective communication. For the time being, brevity should always be secondary to precision and clarity.

It is plain that Internet communication has its own set of rules: rules that do not follow the standards of good English. What is not yet known is whether this is an evolution of the language, or merely a phase that must be endured until the medium evolves, and finds its own voice. Only when a comfortable medium between today's "net-speak" and standard English is reached, can we hope to see any substantial positive changes in the English language.

Words of English through the internet & other technologies.

The design of artificial languages such as Esperanto is commonly regarded as fruitless effort, since they do not tend to gain much popularity. However, the objective need for them is very great and increasing. Computers and information technology have made it possible to process languages automatically. Automatic translation is one aspect of this, and it would benefit from the use of a clearly defined and structured intermediate language. Even more important aspects are the automatic conversion of texts from spoken form to a written one and vice versa and the need for a suitable language that could be used in man-machine interaction. It would be ideal to use a well-designed language (with the expressive power of natural languages) or subsets of it as control languages of computer programs.

Computers can be programmed to process complicated and irregular languages, but the *computational efficiency* is a very important issue, since we wish to process large amounts of text (or speech) and since the processing inherently requires resources which are large even compared with the capabilities of modern computers. For instance, World Wide Web search engines operate on very large and rapidly growing amount of textual information.

A language suitable for automatic processing should be *defined formally* as far as possible. A formal description of the syntax as well as would make it easier to write software for both analyzing and generating the language. An official dictionary should use a rigorously defined formalism to indicate properties of words such as classes of words and transitiveness of verbs. Ideally, a subset of the language itself should act as the metalanguage.

Some features of semantics could be defined formally. This applies in particular to the meanings of derivative suffixes: they should be expressed using a notation which specifies the meaning using an analytic expression.

In compiler technology it is customary to make a clear distinction between *lexical, syntactic, and semantic* analysis. This approach should be applied to the construction of an artificial language: the mapping between letters and phonemes should be independent of other features of the language, such as the grammar and the meanings of the words it should be possible to analyze a sentence grammatically without any lexical knowledge; this is important both for computer processing and for human beings (who can benefit from knowing the grammatical role of a word without knowing its meaning).

Texts are produced, transmitted, and recorded using computers to a rapidly increasing amount. The most common character set is still ASCII, which contains the English letters (A - Z, a - z) and some punctuation and special characters. Although wider character sets have been defined and standardized, general support for them cannot yet be assumed to exist.

Thus, the alphabet for an artificial language should be the English alphabet or a subset of it, without any diacritic marks.

Automatic generation and recognition of speech are technically possible now, and future development will make them economically feasible in a large number of

applications soon. The effort needed to generate and especially to recognize speech strongly depends on the regularity of the phonetic structure of the language.

The following features are desirable:

one-to-one correspondence between phonemes and letters

regular stressing, preferably such that the first syllable is stressed, since this makes it easier to separate words in speech recognition.

The normal principle in modern Indo-European languages is that the vowel of an open stressed syllable is somewhat longer than other vowels, and there is no technical objection to this. It provides additional support for distinguishing the stressed syllable of a word and is easy to generate. However, since the stressed syllable is the first one, this would lead to unfamiliar pronunciation of familiar words in many cases.

Let us first consider the need for grammatical categories such as number, gender, and tenses, without yet considering how they are to be expressed.

Grammatical *gender* is definitely an atavistic feature. It is even undesirable to have to express the natural sex of a being unless it is relevant in the context. For example, there should be a neutral (or more exactly "utral") pronoun that covers both "he" and "she".

Number is astonishingly often an unnecessary or even harmful category. Consider how often one has to say something like "one or more" if one wants to be exact. And consider how often number is specified by other than grammatical methods; in a phrase like "two horses" the plural ending "-s" is of course redundant. Also notice the illogical use of plural in questions like "How many horses have you got?"; the answer might well be "one" or "none", not a plural numeral. We can easily dispense with number as a category. Whenever desired one can use an auxiliary word denoting "one" or "more than one".

Analogous reasoning might lead us to omit *tenses* as well. For instance, in the sentence "yesterday I worked ten hours" the grammatical element "-ed" is as redundant as "-s" because of the adverbial "yesterday". Thus one could well have the rule that the one and only finite form of a verb does not as such specify the time in any way, so that e.g. past tense is expressed by having an adverbial that refers to the past; one should of course have a one-word adverbial that means "in the past" and could be used when a more specific one cannot be used. However, it is useful to have the category of tense for participles, and it would be irregular not to have it for finite forms. More importantly, it would be difficult to express complex temporal relations (e.g. "I will have written") without tenses; using the three natural tenses (past, present, and future) Esperanto allows to express them nicely.

Modes of verbs are hardly needed, in general, since the desired meanings can be expressed using adverbs, different conjunctions for various types of sentences, etc. Modes like subjunctive are difficult to learn and to use in languages like French, and they seldom have any useful meaning. The only exception to "modeless" system of conjugation could be the imperative. The imperative could be used to denote an impersonal instruction or suggestion, as opposite to personal commands, wishes etc. which require delicate distinctions of degrees of politeness and imperativeness, best expressed using adverbs.

Definiteness (as expressed by indefinite and definite articles in many languages) may appear as a necessary category. However, articles are often entirely redundant, and the exact rules for using articles are one of the most difficult features of the English grammar. The following solution is suggested: an indefinite and a definite article exist, but they are to be used only when it is desired to *explicitly designate* a being as "previously unknown" (or indefinite) or as "known" (or definite). In phrases like "the sun" or "the best of all possible worlds" the article is redundant, and in phrases corresponding to "I saw a fox" or "I saw the fox" an article should be

used only if one wants to be explicit about the matter. *Analytic or synthetic* Obviously the language should have a systematic grammatical structure which is easy to recognize and generate automatically. It is not obvious, however, whether this is better achieved by analytic or by synthetic methods.

It is often said that languages develop towards analyticity. But this is probably partially due to the irregularities in synthetic methods. More regularly synthetic languages like the Fenno-Ugrian ones have tended to preserve and even extend synthetic features in the grammars.

From the viewpoint of automatic processability there is no strong objection to synthetic methods. On the contrary, congruence may ease the task of recognizing the grammatical relations between words of a sentence. Moreover, word derivation is essentially synthetic, and the distinction between derivation and inflexion is to a great extent a mere convention.

However, purely synthetic language would be slightly better for automatic processing. This does not apply primarily to computer programs but their human users. It is easier to write a command for searching for a word if one knows that it always occurs in exactly the same spelling.

If synthetic methods are introduced, they must be regular: an affix shall not depend on the base word and shall not affect its form.

A very important aspect is that it should be possible to *"parse" a word into morphemes "mechanically"*, without any semantic or lexical information, solely on the basis of knowledge about the possible affixes. Concretely, base words should not have a beginning or an ending which is the same as a derivation prefix or suffix, respectively. lexical knowledge. Thus, if we have the prefix "re-" (as in Latin and in many international words), then no base word and no other prefix should begin with "re-". This is not easy to achieve. Actually, conflicting prefixes or suffixes like "re-" and "retro-" are not so problematic - we can simply recognize the longest possible

affix. We might allow some conflicts to be resolved by (small) vocabularies of exceptional words (e.g. base words beginning with "re-").

There is, however, a very simple approach which solves the essential problems: use *only one way of attaching affixes* - and suffixing is the obvious approach, since it is more productive in current languages. Thus a word would be morphematically parsed "backwards": simply check if the word ends with a suffix of the language, remove the suffix and apply the same test to the rest of the word etc., until no suffix can be found. Then one can lexically check that the remaining word exists in the basic vocabulary; if not, the word is assumed to be a "foreign" word (a proper noun). Notice that the "suffixing only" principle also removes a semantic problem: assuming that we have a word of the form prefix+base+suffix, is it to be understood as (prefix+base)+suffix or as prefix+(base+suffix)? (Such issues are often real problems. For instance, many people misunderstand the word "atheism" as if it consisted of the negation prefix "a-" and the word "theism"!)

Some very simple and easily definable and recognizable prefixes could be accepted (e.g. the prefix "non-", denoting simple negation of the rest of the word).
Suffixal inflexion and derivation according to the above-mentioned principles requires that each word (excluding such particles which are not used as base words) end in a manner which allows agglutinative suffixing, i.e. suffixing which involves no change in the base word or in the suffix. An obvious way of achieving this is that each word (or, for verbs, the stem of the word) ends with a vowel and a suffix normally consists of one or more consonants followed by a vowel.

It should be as easy as possible to recognize the *constituents* of a sentence, such as subject, predicate, and adverbials. This is often difficult in natural languages, and sentences might be really ambiguous, so that the decomposition into constituents is ultimately based on semantic reasoning. This causes serious problems to nonnative users of the language and even more serious problems to computer programs.

A simple principle which makes the recognition of constituents easier is that whenever there are adjacent nominal clauses (with no prepositions between them) they are *"codeterminative"*, i.e. they form a composite expression which denotes the entity determined by them together. This means that we need not have a distinction between nouns and adjectives. Semantically a word can have a wide range of meanings; when the context is clear, a word can be used as such to denote a specific (context-defined) entity; and when desired, additional codeterminative words can be added to select a particular meaning. Thus, a word might mean "good" both as a noun (a good person or a good thing) and as an adjective.

Notice that the principle of codeterminativeness requires a grammatical sign for the object. In a synthetic language, the sign would be a preposition corresponding to the accusative case. (One might even consider marking the subject, to make the structure more regular.)

We can define the language so that an adverbial which begins with a preposition consists of all nouns following the preposition, up to (and excluding) the end of the sentence, or a preposition, or a finite form of a verb, whichever occurs first. Rules for word order can be set up relatively freely. For automatic generation of sentences, a strict word order would be easier. Languages with "free" word order use variations in the order to indicate nuances. For automatic processing, it would be better to express nuances by using adverbs and by having partial synonyms for important words. Adverbs are definitely to be preferred, since synonyms cause serious problems in search operations: it is difficult to search for texts discussing a phenomenon if there is large set of alternative words for the phenomenon.

From the computational point of view, the most natural word order would be VSO (verb, subject, object), since that corresponds to the normal syntax of subprogram calls. Both the subject and the objects (direct and indirect) are comparable to arguments of a subprogram call, whereas the predicate verb

corresponds to the subprogram name. Adverbials can be regarded as optional arguments, so they should logically appear after other arguments. Notice that normal imperative sentences, which are so common in languages for controlling computers, have the form VSO with the subject omitted.

Word derivation should be extensive to make vocabularies smaller and generally based on suffixes. Each suffix is defined by its actual phonetic (and literal) appearance, its role as deverbal or denominal, its class (noun or verb), and its semantic meaning either as a function of the meaning of the base word or as "to be defined". The latter option means that there is no generic predefined meaning; in that case, the meaning of each word formed using the suffix should be defined separately and listed in a dictionary. (In natural languages, deminutive suffixes are typically morphemes which are thought as indicating smallness only, but in fact they usually belong to the latter category. A cigarette is not a small cigar in reality, just metaphorically.)

Composite phrases consisting of two or more words should be preferred to word composition for the following reasons:

it is difficult to decompose a compound word automatically (and often also by a human reader or listener); this can be alleviated by requiring that a hyphen must be used in compound words, but it is still difficult to recognize such a word from speech

word composition leads to long words which are difficult for human beings.

Thus, word composition should be restricted to cases where a one-word term is obviously needed. It should not be regarded as a fully productive tool, and accepted compound words should be listed in vocabularies. Consequently, the semantics of word composition need not be so well-defined as that of suffixes.

This article has outlined an optimal international language starting from the idea of automatic processability. For further design, it would be necessary to fix such things as the structure of morphemes and the basis of word creation. There have been

several basic strategies of word creation in artificial languages. (Somewhat surprisingly, the one that would most naturally suggest itself in the modern world, using English vocabulary as the basis, is one of the rarest.)

The Loglan language had several design criteria similar to those presented here, probably mainly because they were considered useful for purely human communication as well. It has often been said that Loglan is too logical for human beings to gain popularity, since normal fluent speech does not apply logical forms. On the other hand, Loglan has a feature which is definitely unnatural: its vocabulary is very artificial. It might be a useful experiment to construct a language with structure similar to that of Loglan but with Latin or English based vocabulary.

The advantages of using English through

Generally speaking, when we talk about “language and the Internet”, we are referring to language issues that arise in constructing natural language to be carried across the Internet (e.g., email, chat, the content portion of Web pages). When people talk about “Web languages”, they usually mean coding systems (e.g., HTML, JavaScript) for constructing Web pages. At the intersection of these domains is the problem of figuring out how to design search procedures and coding systems that enable end users to employ natural language input (typically a word or phrase) to locate information posted somewhere in cyberspace on one or more of the billions of Web pages that exist.

Information is of little value if you cannot find what you are looking for. A mislaid book or a document buried in a heap of papers is as good as lost if you need its contents now.

The information retrieval problem is an old one, dating back at least to the great

libraries of Alexandria and Pergamum in the third and second centuries BC. Early classification schemes were often based on the order in which a library acquired a volume – a cumbersome system that was still maintained, in part, by the Library of Congress and the British Library well into the second half of the twentieth century.

Categorization schemes based on subject matter (e.g., the Dewey Decimal call number system, the Library of Congress call number system) would seem to offer better information retrieval. However, Aristotle notwithstanding, there is no "natural" set of categories into which matter or knowledge is divisible. Therefore, library classification schemes (or, these days, choices of key words for finding materials in online databases) are essentially best guesses as to how users are likely to categorize the books or information they are seeking. In a world of small libraries (or bookstores), if the categories made no sense, one could always ask a librarian or proprietor, or simply walk through the stacks. Today, however, we are increasingly dependent upon navigating databases ourselves to find hardcopy books and articles housed in traditional libraries.

Such library challenges are multiplied many hundred-fold as we try to navigate the World Wide Web. Not only is the amount of material being posted to the Web expanding exponentially, but there is no commonly agreed upon set of categories into which that material is sorted. Hence, the retrieval problem becomes truly daunting. Realistically, the only solution for coping with the embarrassment of Web riches is to rely on automatic processing systems that can extract relevant information from natural language text and then, in some cases, categorize, summarize, or even translate such information from one language to another. The question then becomes, are existing natural language processing (NLP) programs adequate for the task? Most of our successful models for doing natural language processing (or machine translation) presuppose that the subject matter, grammar, and vocabulary being processed are constrained. Written texts are assumed to follow basic grammatical

conventions, as well as allowing for a defined set of special conventions (e.g., abbreviations, truncation of full sentences such as “Have you the time?” to “Got the time?”). Moreover, parsers are designed to expect written materials to employ conventional spelling and punctuation.

Natural language usage on the Internet is anything but conventional or constrained. No grammar teacher or subject classifier vets the billions of emails, instant messages, chat contributions, or Web pages that shuttle across the Internet daily. Instead, we find language that is fragmentary, laden with typographical errors, often bereft of punctuation, and sometimes downright incoherent. We also find text dealing with every subject imaginable. If search engines are ever to effectively apply NLP techniques to the breadth of natural language material that appears on the Internet, we will need to understand what sorts of linguistic patterns we are likely to find on the Internet and how they relate to the more conventional written and spoken language that NLP programs now handle.

Since much of the language we find on the Internet is intended as natural language usage (e.g, a posting to a listserv, the home page of corporation), it is important to be aware of the linguistic considerations that come into play when natural language is transmitted via the Internet. The linguistic dimensions we will consider are (1) speech versus writing (2) the effects of technology on language (3) cross-linguistic and cross-cultural issues. Most of the issues surrounding natural language usage on the Internet are not unique to the medium. Rather, the variables shaping such usage are familiar from studies in linguistics concerning transmission modality (here, speech versus writing), discourse participants, style, transmission medium (e.g., spoken language delivered face-to-face vs. over the telephone) and culture. An understanding of these issues is important for anyone hoping to tackle the challenges of creating Web search tools that work successfully with the natural language appearing on the Web. The overwhelming majority of natural language

appearing on the Internet is written (i.e., as opposed to auditory). However, Internet users have often commented that written language on the Internet, especially in email or now instant messaging (IM), is more like speech than like writing. Before analyzing such comments (much less assessing their veracity), we need to understand the relationship between speech and writing as modalities through which to formulate and convey human language.

Technological developments have long affected the nature and use of written language. The establishment of print technology eventually fostered not only standardization of spelling but the growth of literacy in Western Europe. Successful marketing of typewriters at the end of the nineteenth century fueled invention of the business memorandum, contributed to the decline of American handwriting skills, and encouraged the production of more lengthy typewritten prose, a trend also evident in the early years of word processing.

There are approximately 5000 different languages in the world today. Historically, there have been four basic approaches to the challenge of communicating across these linguistic boundaries: (1) remaining monolingual (i.e., not communicating across linguistic boundaries) (2) multilingualism (i.e., learning the language of another language community, such as German speakers learning French) (3) creation of a contact language (e.g., a pidgin or creole, such as Tok Pisin in Papua New Guinea or Krio in Sierra Leone) (4) adoption of a lingua franca (i.e., a common language that may be the native language of few or none of the specific discourse participants, such as historically happened with Latin and French, and is now happening with English). While some non-native speakers may become fluent in the fullfledged version of the lingua franca, others may function with a restricted subset of the language (e.g., Airspeak for international aviation, the somewhat simplified version of English used by Voice of America, and what some describe as an emerging simplified version of English used by millions of non-native speakers in

communicating across linguistic boundaries) Since the Internet is a global system, issues relating to linguistic diversity are critical in determining how to deal with natural language carried via the Net. Although the Internet developed as an English-based network, the use of other languages has rapidly been expanding. There is the well-known projection (appearing in a 2001 airport billboard advertising Accenture Consulting) that by 2007, the dominant language on the Internet will be Chinese. This projection is hardly surprising in light of the estimate that by the year 2050, there will be nearly 1.4 billion native speakers of Chinese but only slightly over 500 million native speakers of English.

The second trend is towards adoption of English as the lingua franca of the Internet. However, two issues arise with this solution. On the one hand, since the majority of English speakers in the world already are non-native users of the language, content writers for the Internet cannot assume that even the majority of readers will understand complex grammatical constructions, idioms, or less common vocabulary.

On the other hand, even among native speakers (or speakers of English-based pidgins or creoles), there remains the question of which dialect of English to select.

The famous Peter Steiner cartoon that appeared in the *New Yorker* in 1993 had one canine confiding to another, “On the Internet nobody knows you’re a dog”. As of now, the overwhelming majority of natural language text carried across the Internet is written, thereby neutralizing differences in the spoken-language accent structure of Internet users from, say, Dallas, Dublin, and Delhi. However, other tell-tale differences between English dialects reveal themselves in writing: through spelling, vocabulary, and occasionally grammar. Is the word spelled *color* (American English) or *colour* (British English)? Does a mail-order company *send* a product or *dispatch* it? When you are sick, do you go “to the hospital” or “to hospital”?

Search engines and parsers are unlikely to have much difficulty handling differences between the major dialects of English. However, members of distinct cultural communities may be highly sensitive to linguistic nuances. A single spelling of the word *colour* immediately signals "foreign" to an American reading the word on the Internet.

Among the culturally-based issues of writing style that arise with natural language usage on the Internet are these:

► **audience**

Are there people to whom it is inappropriate to send particular messages (e.g., a freshman biology student emailing a Nobel prize winner for help with her homework)?

► **content**

Are there some messages that should not be sent over the Internet (e.g., condolences) but only in a formal off-line letter? Are there some messages that should only be delivered face-to-face (e.g., dissolving a personal relationship)?

► **form**

How formal or informal should the style be (e.g., should you use a salutation in email, when should you use first name versus title plus last name)? Is the message edited before it is sent? (In some cultural contexts, unedited messages are perceived as rude.)

In 1999, a Montreal photographer named Michael Calomiris was fined by the Quebec government because of the content of the Web site he used to display his works. The issue was not the photographs but the language of the site. By Quebec law (Section 52 of the Charter of the French Language), it is illegal in Quebec not to do business in French. Calomiris's site was in English because, as he explained, "my aim was global, not local….We have a lot of customers in the United States, and …they don't care if [the site is in] French".

There is a well-known story, though apparently apocryphal (Hutchins 1995), about early problems in using computers to translate between languages. A generic version of the story goes like this: Soon after the onset of the Cold War, scientists were interested in testing whether computers could successfully translate between English and Russian, given the importance of accessing Russian documents. Choosing a Biblical text (Matthew 26:41), the computer operator fed in the English words "The spirit is willing but the flesh is weak" to produce a Russian translation. The Russian result was then fed back into a Russian-English translator program, yielding "The vodka is strong but the meat is rotten" – hardly a result that might bolster national security.

Beyond the challenge of dealing with thousands of different world languages, there is the problem of handling the dozens of scripts used to encode those languages. Linguists commonly divide scripts into three categories. The first, logographic, uses symbols to stand for whole concepts or words. Chinese is the best known example. The other two basic script types represent sounds. Syllabic scripts (e.g., Japanese *hiragana* and *katakana*) use a single symbol to stand for an entire syllable (e.g., the sound sequence *ka* in the word *katakana*). Alphabetic scripts (e.g., Greek, Arabic, and Roman) pair individual sounds with individual letters (e.g., representing the sound sequence in the word "cat" with three distinct graphemes, {c}, {a}, and {t}). However, the situation with scripts quickly becomes far more complex. Within a single script type (e.g., alphabetic), there is considerable variation. Greek, Russian, Arabic, and English are all written with alphabets, but with different alphabets, due to the different historical circumstances under which their writing systems arose. A second complication is that even within a given script subset (e.g., the Roman alphabet), there is variation across languages using that written script. For example, written French uses several diacritical marks (e.g., *accent grave,* as in *è*; *cedilla,* i.e., *ç*) not found in English. Written Polish has nine alphabetic symbols (written by

adding diacritical marks to existing Roman graphemes – e.g., *ł* and *ź*) that are not found in written English. These slight alphabetic variants can prove disastrous for contemporary search engines, since no common conventions have been established for how search engines handle variants of Roman script.

Finally, there is the linguistic and political reality that the script used to encode a language may change over time, necessitating doing Web searches of the same language in multiple scripts. For example, traditional Chinese characters are used in Taiwan, while simplified Chinese characters are used in the People's Republic of China. Historically, Turkish was written in an Arabic-based script, though with the reforms of Kemal Ataturk in the early twentieth century, Turkish came to be written in a modified version of Roman script.

A growing number of countries that employ scripts other than the basic Roman character set used for English have developed keyboards and computer software enabling writers of, say, Arabic, Chinese, or Hebrew to do word processing, send email, and engage in native-language Internet activity. However, a problem arises when users of different scripts (or script variants) attempt to communicate with one another across the Internet .

Conclusion

The World Wide Web, in particular, offers a home to virtually all the styles which have so far developed in the written language: newspapers, scientific reports, bulletins, novels, poems, prayers - you name it, you'll find a page on it. Indeed, it's introducing us to new styles of written expressions which none of us have ever seen before - animated language in particular. Words which appear and disappear in varying colours; sentences which slide onto the screen and off again; letters which dance before your eyes. The Web is truly part of a new linguistic medium, more dynamic than traditional writing, and more permanent than traditional speech. It's often been said the Internet is a revolution - yes, indeed, but it's also a linguistic revolution.I can give you a couple of quick examples, so that you see my point. Take emails. You receive a message which contains, say, three different points in a single paragraph. You can, if you want, reply to each of these points by taking the paragraph, splitting it up into three parts, and then responding to each part separately, so that the message you send back then looks a bit like a play dialogue. Then, your sender can do the same thing to your responses, and when you get the message back, you see his replies to your replies. You can then send the lot on to someone else for further comments, and when it comes back, there are now three voices present on the screen. And so it can go on, replies within replies within replies. And all looking exactly the same, in the same screen typography. There's never been anything like this in the history of human letter-writing. That's one reason why I say the Internet is a linguistic revolution.Or say you're participating in a real-time Internet discussion group -a chatroom. You see on your screen messages coming in from all over the world. If there are 30 people in the room, then you could be seeing 30 different messages, all making various contributions to the theme, but often clustering into half a dozen or more sub-conservations. It's like being in a cocktail party where there are

other conversations going on all around you. In the party, of course, you can't pay attention to them; in a chatroom you can't avoid them. It has never been possible before, in the history of human communication, to attend to 30 people all talking at once. Now you can. It's a revolution, all right.

But there's another reason for the revolutionary status of the Net - and this one may surprise you. It's because the Internet offers a home to all languages - as soon as their communities have a functioning computer technology, I mean. Its increasingly multilingual character has been the most notable change since it started out - not very long ago - as a totally English medium. There's a story the former US Vice President, Al Gore tells. He was reporting the remark of the eight-year-old son of Kyrgyzstan's President Akayev, who told his father that he had to learn English. When asked why, the child apparently replied, 'Because, daddy, the computer speaks English'.

For many, indeed, the language of the Internet is English. There was a headline in The New York Times in 1996 which said simply, 'World, Wide, Web: 3 English Words'. The article, by Michael Specter, went on to say: 'if you want to take full advantage of the Internet there is only one real way to do it: learn English'. He did acknowledge the arrival of other languages: 'As the Web grows', he said, 'the number of people on it who speak French, say, or Russian, will become more varied and that variety will be expressed on the Web. That is why it is a fundamentally democratic technology', he said, 'but it won't necessarily happen soon.'

Well the evidence is growing that this conclusion was wrong. With the Internet's globalisation, the presence of other languages has steadily risen. By the mid-1990s, a widely quoted figure was that about 80% of the Net was in English. People were acknowledging the first major study of language distribution on the Internet, carried out in 1997 by an organisation called, believe it or not, Babel, a joint initiative of the Internet Society and Alis Technologies. This showed English well

ahead, but with several other languages entering the ring, notably German, Japanese, French and Spanish.

Since then, the estimates for English have been steadily falling. Some commentators are now predicting that before long the Web, and the Internet as a whole, will be predominantly non-English, as communications infrastructure develops in Europe, Asia, Africa and South America. Listen to these results of a recent Global Reach survey. They estimated that people with Internet access in non-English speaking countries increased between 1995 and 2000 from 7 million to an amazing 136 million. In 1998, there was another surprise: the number of newly created Websites not in English passed the total for newly created sites that were in English. And at a conference on Search Engine Strategies in London last year, Alta Vista were predicting that by 2002 less than 50% of the Web would be n English. In certain parts of the world, the local language is already dominant. According to one Japanese Internet author, Yoshi Mikami, 90% of Web pages in Japan are already in Japanese. The Web is increasingly reflecting the distribution of language presence in the real world, and many sites provide the evidence. There are thousands of businesses now doing their best to present a multilingual identity. For instance, the Belgian daily newspaper, Le Soir, is represented by no less than six languages, French, Dutch, English, German, Italian, and Spanish. Then there are now hundreds of major sites collecting all kinds of data on the languages themselves. Call up the font archive at the University of Oregon, for example: you'll find there 112 printing fonts in their archives for over 40 languages. They have a nice sense of humour, too - because you'll also find some data there on alien languages, such as Klingon, and folklore languages, such as Elvish, which Tolkien invented for Lord of the Rings. Spend an hour hunting for languages on the World Wide Web and you'll find hundreds. I found one site, called World Language Resources, which lists products for 728 languages. I found an African resource list which covered several local

languages: Yoruba, for example, was illustrated by some 5,000 words, along with proverbs, naming patterns, and greetings. Another site dealt with no less than 87 European minority languages. Some of the sites were very small in content, of course, but nonetheless extensive in range: one gave The Lord's Prayer in nearly 500 languages.

Nobody has yet worked out just how many languages have obtained a modicum of presence on the Web. I found over 1,000 quite quickly. It's not difficult to find evidence of a Net presence for all the more frequently used languages in the world, and for a large number of minority languages, too. I'd guess that about a quarter of the world's languages - that's about 1500 - have some sort of cyber existence now.

On the other hand, I have to recognise that developing a significant cyber-presence for a language isn't easy. There's a sort of 'critical mass' of Internet penetration which has to build up in a country before a language develops a vibrant cyber life. It's not much use, really, to have just one or two sites in a local language in the Web. People wanting to use or find out about the language would soon get bored. The number of sites has to build up until, suddenly, everybody's using them and adding to them and talking about them. That's a magic moment, and only a few hundred languages have so far reached it. In the jargon of the Internet, there needs to be lots of good 'content' in the local languages out there, and until there is, people will stay using the languages that have managed to accumulate content - English, in particular.

So the future of a multilingual Internet isn't guaranteed. It will all depend on how quickly new sites can build up a local language momentum. There are also a number of practical difficulties. Until quite recently there were real problems in using the characters of the keyboard to cope with the alphabetical diversity of the world's languages. Because it was the English alphabet that was the standard, only a very few non-English accents and diacritics could be handled. If it was a foreign word with

some strange looking accent marks, the Internet software would simply ignore them, and assume they weren't important. This can still happen - but things have moved on a great deal since then. First, the basic set of keyboard characters, the so-called ASCII set, was extended, so that the commoner non-English diacritics could be included. But even then it only allowed up to 256 characters, and there are far more letter shapes in the world than that. Just think of the array of shapes you find in Arabic, Hindi, Chinese, Korean and the many other languages which don't use the Latin alphabet. Today, a new coding system, the UNICODE system, is much more sophisticated: it allows the representation on screen of over 65,000 characters. That should be plenty - but the implementation of this system is still in its infancy.

My feeling is that the future looks good for Web multilingualism. Web offers a World Wide Welcome for global linguistic diversity.

Bibliography

1. И.А.Каримов. “Чет тилларни ўрганиш тизимини янада такомиллаштириш чора тадбирлари тўғрисида”ги қарори.”Ҳалқ сўзи “газетаси .11.12.2012й. 240(5660)-сон)
2. Andrews, Jean. 2002. *i-Net+ Guide to the Internet*, second edition. Boston: Thompson Learning. 145-145pp
3. Baron, Naomi S. 1981. *Speech, Writing, and Sign*. Bloomington, IN: Indiana University Press.211-213pp
4. Baron, Naomi S. 1986. *Computer Languages: A Guide for the Perplexed*. New York: Doubleday.254-256pp
5. Baron, Naomi S. 1998. Letters by Phone or Speech by Other Means: The Linguistics of Email. *Language and Communication* 18:133-170.
6. Baron, Naomi S. 2000. *Alphabet to Email: How Written English Evolved and Where It’s Heading*. Routledge: London.
7. Berners-Lee, Tim. 2000. *Weaving the Web*: New York: HarperCollins.
8. Berners-Lee, Tim, James Hendler, and Ora Lassila. 2001. The Semantic Web. *Scientific American*, May, 35-43.
9. Biber, Douglas. 1988. *Variation across Speech and Writing*. Cambridge: Cambridge University Press.
10. Bloomfield, Leonard. 1933. *Language*. New York: Holt.
11. Bolter, Jay David. 2001. *Writing Space*, 2nd edition. Mahway, NJ: Lawrence Erlbaum.