

The Importance of 'International' Words to Auxiliary Languages and How to Define Them

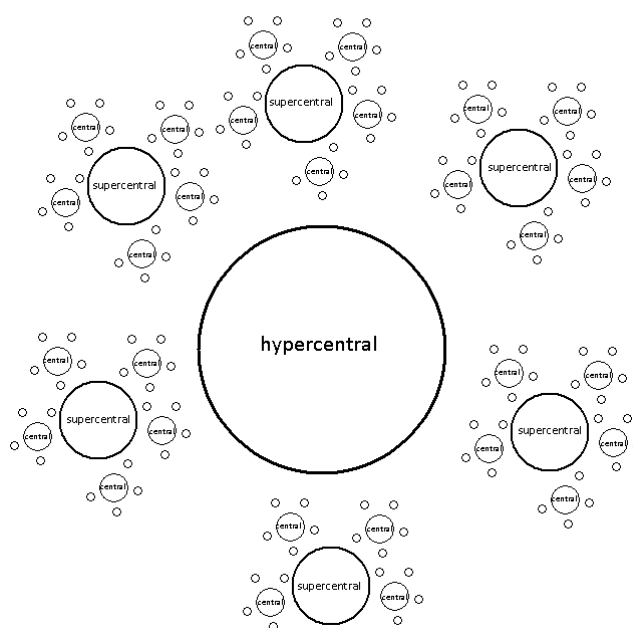
Many international auxiliary languages (IALs) are judged based on the familiarity of their vocabulary to speakers of various languages around the world. But which words are suitable to be borrowed in - and from which languages? In this talk, Johan discusses criteria for incorporating loan words into IALs that have been used throughout history, and also proposes criteria of his own.

The world languages system, shown here, is a concept introduced by Abram de Swaan. Languages form a coherent whole. They interact with one another, compete with one another and are engaged in a power struggle.

Of course this struggle is not a planned one, although in 1943 Winston Churchill advocated Basic English as a world tongue by saying: "Here you have a very carefully wrought plan for an international language, capable of very wide transactions of practical business and interchange of ideas." and undoubtedly the State Department in Washington is very much interested in the significance of promoting the English language and the economic advantages for its international commerce.

This unplanned struggle takes several forms: marginalizing and displacing languages, up to complete devouring them.

More peaceful intrusion is the so called loaning of words. No language can escape the thrust of foreign words coming with new commodities, techniques, social methods, science and youth culture, although the attitude of language communities against this thrust is very diverse. Icelandic and German say *Sjónvarp* and *Fernsehen* in stead of something close to television. French is much more subject to intrusion of English words than is generally assumed. And the French government adopted strong legislation to curb the replacement of originally French words and inventing new ones like *courriel* for e-mail and *logiciel* for software.



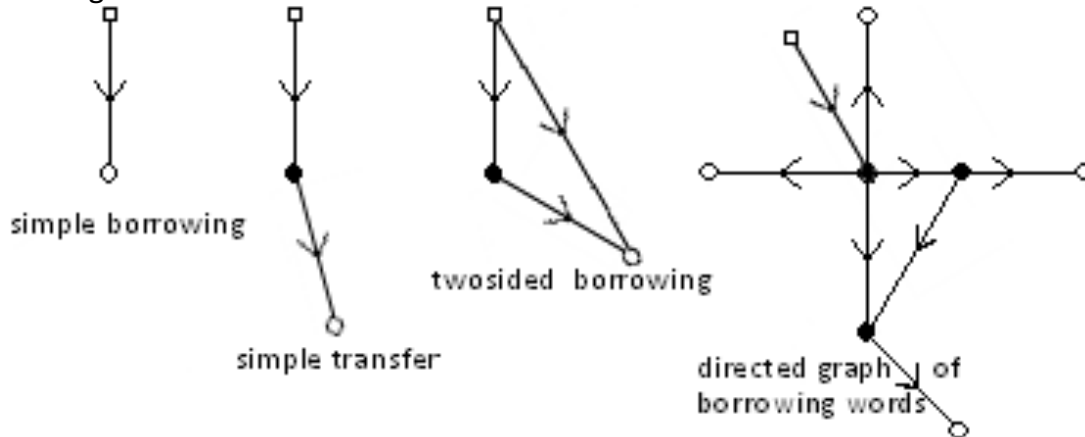
The model of competing languages by Bram de Swaan categorizes languages as 'periferal', 'central', 'supercentral' and 'hypercentral'. The one hypercentral language is English, supercentral are Arabic, Chinese, French, German, Hindi, Japanese, Malaysian, Portugese, Russian and Spanish. National languages are most often 'central', while periferal languages often are not much written and are not standardized.

The structure of dependency of languages defines the direction in which marginalisation and loaning of words takes place: top – down. English lends to German, German lends to Turkish, Turkish lends to Kurdish, not vice versa. This lending of words however depends on the linkages people of one language have with another language. Many Turks went to work in Germany and Kurds went to Turkish cities.

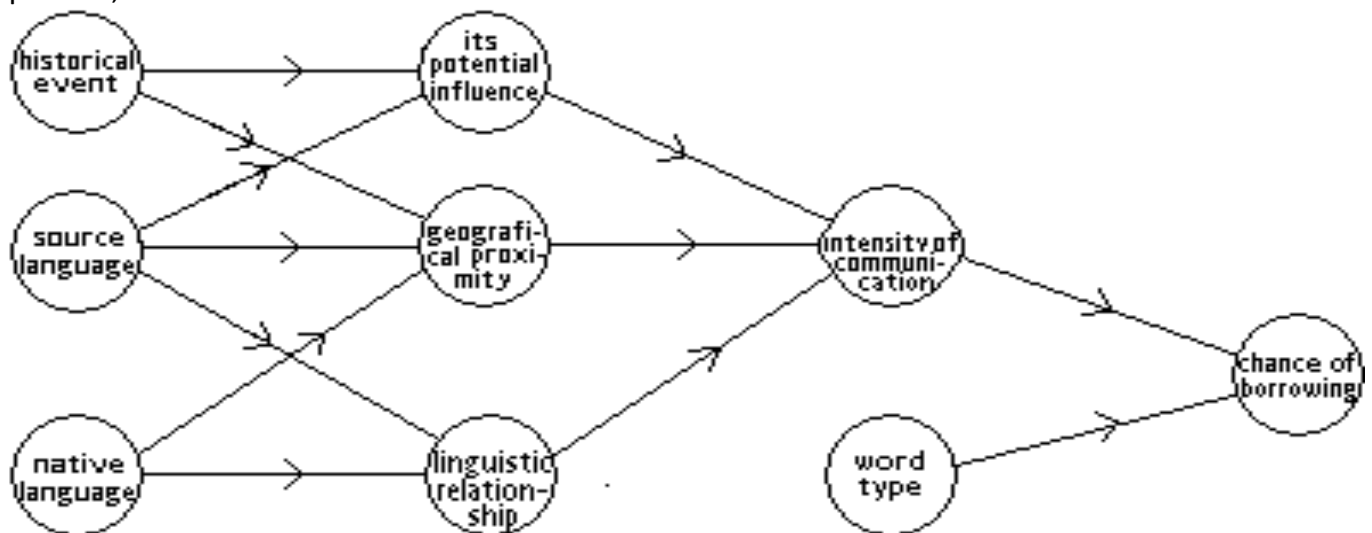
People from different languages tend to communicate in a language to which both are linked higher in the hierarchy. So Flemish people often communicate with Walloons in English, although French has a higher status than Dutch, but those two languages are not closely linked.

The difficulties of this rationalized scheme are numerous, mainly because it does not include political and cultural differences and the function of identification a language has.

On an individual level, the level of single words or expressions, one can think of many patterns of loaning:



Focussing now on the borrowing of a specific word one can distinguish many factors influencing the process, for instance:



At the middle left you see the source language, which exerts its influence (top center) in a certain historical period on other languages, native languages. Depending on that period the source language has a certain extension over the world and has contact by way of the available means of communication with a native language (i.e. geographical proximity), not to speak of warfare and colonisation. Also a factor pertaining to the intensity of communication is the linguistic relationship between both languages. Finally the borrowing process may be caused and supported by production and trading of objects or phenomena, whose names are new to the speakers of the native language. This causes different types of word.

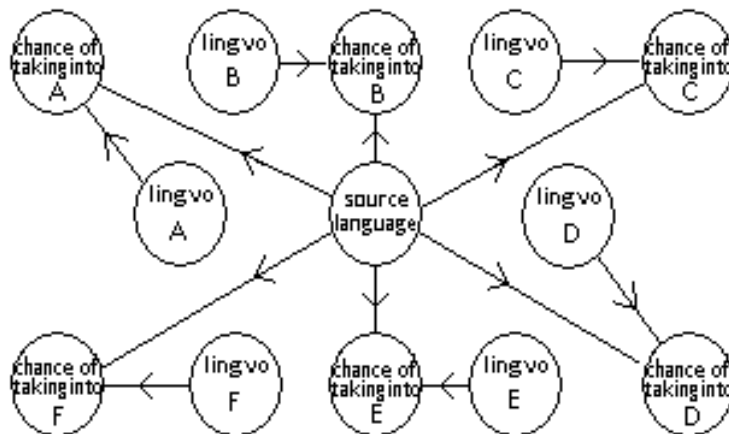
You all know from recent European history how words in certain economic branches were imported from one language and words for cultural activities from another: musical terms and banking from Italian, cuisine and clothing fashion from French, technical words from German, shipbuilding from Dutch, sport terms from English.

Because all those influences follow each other, each language becomes a multilayered structure of foreign influences. To create a model of this multipolar world then it is necessary to abstract from historical period.

At first we simplify the flowchart just shown and have



But one powerful language like English lends the same words to many languages. So another diagram to account for six borrowing languages is:



What is the relevance of creating such a flowchart for the business of conlangs ?

If we consider the three artificial languages Interlingua, Ido and Esperanto, they represent with respect to borrowing words from natural languages a small spectrum.

After World War two Interlingua has begun to rely for its vocabulary more and more on the great European languages. The passive understanding of Interlingua rests completely on the principle of loan words. The developers of Ido in the twenties of the last century screened many words who they thought were feasible for adoption just on their 'international' intelligibility.

And also for Esperanto applies that the more international a word is, the more fit it seems to adoption.

The first grammar and vocabulary were edited by Zamenhof in English, French, German, Polish and Russian. Apart from Polish those are the main source languages, to who should be added still Italian and Latin.

At the end of the nineteenth century these six living languages constituted in European eyes the civilised world.

Zamenhof wisely incorporated into the Esperantogrammar a rule which states:

The so-called "foreign" words, i.e. words which the greater number of languages have derived from the same source, undergo no change in the international language, beyond conforming to its system of orthography.

So the foreign word must be transferred by tge greater number of language to Esperanto.

The difficulty with this rule is that "the greater number of languages" can hardly be attained. If there would be in the world just one cultural community with one great cultural ancestor like often the Greek civilisation for the Western world, there would be no problem. There is a great number of European words derived from Greek, but in a world with about eleven supercentral languages of

which five are non-western (Arabic, Chinese, Hindi, Japanese and Malaysian) each of them constitutes a word loan center of its own.

Or has the English language as the main medium of western culture and science, already played such a unifying role that people from the most diverse cultural backgrounds have a sizable number of 'international' words in common ?

In order to investigate this question I made up a sample of seventy fairly recent loanwords from English (after World War two up to 1990) and registered their acceptance or rejecting in 54 languages out of Asia, Africa and Europe.

Just to give you an impression of the kind of words I will mention some of them, alfabetically:

aerobics, bazooka, bingo, bodybuilding, booby trap, bulldozer, camcorder, camper, cashew nut, computer, cornflakes, cracker, cryptogram, dinky toy, donut, doping, dressing, drug, email, fast food, football pool, frisbee, hot dog etc.

With the help of 'Google translate' I noted the 54 translations of those 70 words and checked wether they could be considered as being derived from the English word. The results of this check constitute for each language a pattern of acceptance over those 70 words and these patterns can be compared between languages.

(It is reasonable to suppose that central and supercentral languages who are both in contact with the hypercentral language English and have no other common supercentral language from which to borrow a word are both subject to borrowing from English.)

Then it is possible to calculate the degree of conformity between them, conformity in the art of accepting a word from English or, of course, rejecting it.

The most ready to hand method then to group languages on the basis of this degree of conformity is called average linkage clustering. Each language is grouped with its closest match in borrowing behaviour towards English and later languages are added to existing groups and groups to groups untill all of the languages form one single group.

The view of this picture may surprise you. **(See: tree of conformity)**

For instance Greek and Hungarian have 72% conformity, Turkish finds itself in the midst of Romanic languages, Swahili and Maltese are in the midst of Germanic languages, which is not so exceptional because both language are in very close contact with English. And also Japanese positions itself almost as a Western language. Romanic languages are not so close to each other in borrowing pattern as one would think: Italian, Rumanian and French behave apart from the Iberian language group. Have Tagalog, the Philippine language, and Korean something in common ? Is Azerbaijan close to Malayan and Indonesian because of its islamic strain ? Clearly Mandarene and Arabic find each other in a common pattern of rejection of English words.

This tree is not very revealing though. The main finding is that languages in general do not position themselves to twentieth century English words accurately according to their language family.

What more decisive language groups can one discern then in this respect ? Therefore I repeated the process of composing language groups according to their borrowing behaviour, but stopped extending them when the average degree of conformity of the group to the languages outside the group rose above the average degree of conformity inside the group. It appeared that some languages, especially those with a strong identity of their own were difficult to classify with others, while the group most integrated and subject to English influence is of course the Germanic language group with Swahili, Japanese and Georgian and two Celtic languages included.

After the Germanic language group follows in order of declining integration:

(See: Degree of Conformity in word-borrowing of post-WW-II English words of and in between for this purpose best fitted Language Groups)

2. the group of formerly British colonized countries with languages Urdu, Korean, Hindi and Tagalog.
3. a basicly Romanic group, including Basque, Turkish, Hungarian and Greek
4. the Balto-Slavic group with eighteen members to which belong also Finnish, Estonian and Rumanian.

5. a group consisting of Armenian, Thai, Arabic and Mandarene, which may be marked as non colonised Afroasiatic
6. a so called Islamic-like group with Indonesian, Malayan, Azerbaijan, Vietnamese and Persian

We treated up to now only with words in English, the hypercentral language, but there are also lots of words borrowed recently or in earlier ages from supercentral languages and from other central languages with whom existed intensive contact. I have a feeling that analysis of samples of potential loanwords from supercentral languages may reveal a pattern not essentially different from the one just shown, because of several reasons:

- part of the words taken from English are just transfers of words from French, Greek or Latin
- the cultural attitude to other European languages than English of language communities with a strong identity of their own is basically the same

Only languages which entered into close contact with English because of colonisation like Swahili, Hindi, Tagalog, Korean and Japanese supposedly borrowed only a few words from French and German. Also many languages in Asia took many words from Arabic and Turkish, while Spanish greatly influenced central and periferal languages in South-America.

But I will submit to you now an extra reason why the borrowing from supercentral languages may not greatly influence the clustering of languages evolved on the basis of these seventy English words. Therefore I just recall the maxim of the Esperanto Grammar which states:

The so-called "foreign" words, i.e. words which the greater number of languages have derived from the same source, undergo no change in the international language, beyond conforming to its system of orthography.

We only considered one source and even then it is hardly possible to reach a majority of languages, their number being about six thousand. Nor if we take the number of native speakers of all languages into account, because a majority for a word without it being common in Mandarene would be impossible.

The easy way to esperantists out of this dilemma is to point to the fact that Zamenhof thought only of six or seven languages and that we can't without serious internal struggles change this law of Zamenhof.

I advocate the view that a conlang which aspires to be spoken all over the world must accept that only the educated are prone to get interested in the special community which supports this conlang. There are no periferal languages in which you can follow a study at university. In bilingual countries with one European language definitely most of the universities offer courses only in that language. Think of South Africa with twelve official languages, India with twenty six and all other colonies, wether speaking French, Spanish or Portuguese.

A sensible interpretation of the rule of Zamenof can only be like:

The so-called "foreign" words, i.e. words which the greater number of languages, weighed according to each number of entrants of Academic courses, have derived from the same source etc.

I don't want to discuss about the opportunity of just this criterion. Others like Abitur, A-certificate level are of course possible. My choice is based simply on the fact that on the internet the numbers of entrants at university level can relatively easy be found.

English	35.488.428	Vjetnamese	1.654.846	Serbian	237.598
Mandarene	26.924.620	Tagalog	1.325.733	Danish	232.194
Spanish	14.130.662	Rumanian	1.200.223	Slovak	229.477
Russian	13.522.477	Bengal	1.145.401	Norvegian	212.672
Hindi	7.431.481	Malayan	857.598	Lithuanian	204.767
Arabic	6.792.200	Dutch	812.826	Basque	178.102
Portugese	6.417.351	Svedish	561.703	Hebrew	162.623
Indonesian	4.419.577	Birmese	507.660	Finnish	154.824

French	3.980.339	Urdu	486.896	Mongolian	151.533
Japanese	3.938.632	Catalan	534.765	Uzbek	149.505
Persian	3.405.676	Hungarian	413.715	Kirghiz	148.134
Korean	3.204.310	Greek	400.703	Azerbaijan	141.896
Ukrainian	2.847.713	Czech	394.544	Croatian	139.996
Turkish	2.545.466	Kazakh	359.901	Amharic	132.411
German	2.520.962	Galician	356.204	Georgian	129.926
Thai	2.417.262	Belorussian	288.340	Latvian	127.760
Polish	2.146.926	Bulgarian	264.463	Kmer	122.633
Italian	1.882.382	Nepalese	255.354	Slovenian	115.445

and still thirty languages more with numbers of students at third level down to 73 for the language of Maldives.

We are ready now to test the internationality of words in Interlingua, Ido or Esperanto on the basis of my interpretation of the Zamenhofian criterion.

I did so with 95 words, which are listed as 'not recommended' on a website of 'la bona lingvo', which means the righteous language. This site is administered by a group of esperantists who are eager to stress the need for Esperanto words being simple. They generally do not accept loanwords, if a schematic synonym, i.e. a word composed of two or more primitives is available. However all of these words were either declared official by the Esperanto Academy or were in use by Zamenhof himself, the initiator of Esperanto.

It turned out that 39 words out of those 95 were deemed to be understood by more than half of the worlds student population at university level and as such should be a natural part of Esperanto. They have been order on the list you see now (**See: 'international' words**) by declining degree of internationality.

On the next graph (**See: ordigo de vortoj lau nombro da studentoj enprenovaj**) you see all 95 words and the legend in Esperanto. The colours show to which degree each language group has taken in the concerning word. The 56 words at the right of this graph command less than half of the worlds student population and they were rightly not recommended on the website of 'la bona lingvo. Have a look at them (**See: rightly not recommended words**)

For clarity I must add that the 39 international words can't be accepted as such, because the rule of Zamenhof says also that the foreign words should be changed "conforming to the Esperanto system of orthography." and further

"Such is the rule with regard to primary words, derivatives are better formed (from the primary word) according to the rules of the international grammar."

I hope you enjoyed in getting to know some things about which hard-core esperantists strive among themselves.

Thank you very much for your attentive listening.