MULTILINGUALISM AND THE WEB

Elisabeth Burr
Universität Bremen

Abstract: Only a few years ago, the new technologies in general and the web in particular seemed to constitute a danger for most languages of the world. In the meantime, however, it would seem that quite the opposite is the case and that the new technologies favor multilingualism, the deconstruction of norms and of linguistic institutions, and the formation of hybrid languages. In fact, not only has the web itself become a multilingual environment but also minority languages and local dialects are migrating into this digital space. Furthermore, the (intentional) development of mixed languages even in environments dominated by writing would seem an indication of the weakening of the national and standard language paradigm, of the search for a *we-code* counterbalancing institutionalised *they-codes* and for a linguistic identity built on the multilingual knowledge of the speakers. In this contribution I arrive at the conclusion that the process from monolingualism to multilingualism and to hybridity as an important ingredient of digital culture a) presents linguistics with a unique possibility to study directly how the ‘oral’ is being transformed into the ‘written’, how norms evolve and spread, and from which tools depends the recognition of languages or dialects as modern communication systems, b) questions those schools of linguistics which have modelled language as a homogeneous entity, the ideal speaker-listener’s competence as inherently monolingual, and have reduced linguistic variety to a common core, and c) constitutes a serious challenge for Humanities Computing and the developers of language technologies which, if taken up, will open up, finally, this new domain of research to the many Creoles and mixed languages spoken (and written) in the world.

Parole chiave: (hybridity, multilingualism, mixed languages, new technologies, humanities computing)

1 Introduction

Only a few years ago, the new technologies in general and the web in particular seemed to constitute a danger for most languages of the world. The parallels between

- the invention of the printing press, which led to the normalisation of languages for economic reasons and caused the disappearance of language varieties and dialects, and the diffusion of the new technologies,
- the creation of hegemonic centralised European national powers, which forced the imposition of monolingualism, and the expanding hegemonic position of the US,
• the concept of *bon usage* or standard language much favored by nation states and by their educational systems, and the concept of *centre* and *periphery* or homogeneous language communities favored by modern linguistics,

seemed, to indicate, in fact, that another detrimental process of linguistic homogenisation and normalisation was to be expected.

In the meantime, however, it would seem that quite the opposite is the case and that not monolingualism, homogenisation and normalisation are favored by the new technologies but multilingualism, the deconstruction of norms and of linguistic institutions and the formation of hybrid languages.

2 From Monolingualism to Multilingualism

The challenge which multilingualism constitutes in the digital world has not just been recognized by firms like Microsoft or Netscape which have developed and continue to develop language and culture sensitive versions of their browsers, but also by online marketing communications consultancies like the US based GlobalReach, whose aim it is to facilitate the access of transatlantic firms to the European market and who, as the following quotes from their web site show, are extremely conscious of the economic importance of languages in the digital market:

Online marketing always takes place in the language of the target country, as any form of marketing does. (GlobalReach 2004d).

People surf the Web in their own language: it is as much a part of their life in their own language as other media (TV, radio, printed publications, etc.). (GlobalReach 2004e)

Web marketing is global by nature, and requires a multi-lingual approach, since each market has to be addressed in its own language. (GlobalReach 2004e).

The rising market value of languages in the digital world is, however, only one of the aspects of the growing tendency towards multilingualism. Another aspect is the clear indication to be found on the web that the traditional view of multilingualism as representing a burden for humankind is substituted with the consciousness that multilingualism in itself constitutes a cultural value. Emblematic for this change in perspective is the story and the image of the Tower of Babel.

In fact, whereas the story of the Tower of Babel during the Middle Ages was used to explain linguistic diversity as being the result of the linguistic confusion God inflicted on humanity for questioning the divine hierarchies, and whereas later, during the phase of linguistic normalisation brought about by the advent of the printing press and the creation of national states, the same story was regularly exploited by grammarians when justifying their proscription of variety and their establishment of a unique lexical and grammatical norm, in the last decade, on the contrary, Babel and Pieter Bruegel’s painting of the building of its tower has become one of the most frequently used symbols on the web when multilingualism, seen in terms of cultural richness, is at stake.
Furthermore, forecasts like

No speak english nella Rete del 2000 (Mola 1999)
La Babele delle lingue sopravvive in Rete (Retico 2000)

have long become true and the web which, when invented in 1993 and when first known above all in the Anglo-American world and in regions like New Zealand, Australia and Canada was a truly monolingual, i.e. English resource, has itself become a multilingual environment.

2.1 National languages

As the statistics of GlobalReach which has been systematically tracking the development of the online populations since 1996, show, in fact, since 1995 when the first non-English speaking countries connected to the Web, the percentage of the native English speaking online population has gone down continuously: from 84% in 1997 to 64.9% in 1999, to 47.5% in 2001, to 35.6 in September 2003. At present the relationship between native speakers of English and native speakers of other languages on the Web seems to have stabilized at around 35.2% to 64.8% respectively.
During the same period the percentage of English Web pages went down from about 100% to 68.4% (cfr. GlobalReach 2004c)\(^1\) and the number of languages on the Web has grown explosively.

2.2 Minority languages and dialects

Multilingualism on the Web does not mean, however, the presence of national languages alone. Instead, there are clear indications that living minority languages and dialects are also finding their way into the Web, a phenomenon which Italian newspapers thought to be worth reporting already quite a few years ago when articles like the following appeared:

- Altro che inglese: Internet ora parla napoletano (Ferraiuolo 1998))
- Proverbi e aforismi anche in dialetto (Mandò 1998b)
- Minoranze linguistiche e dialetti in Rete (Mandò 1998a)
- I dialetti italiani sopravvivono online (Benigni 2000)

As newspaper articles as such are not a reliable enough proof that dialects and minority languages are really migrating into the Web, research into the presence of northern Italian dialects and minority languages in the digital world has been undertaken (cfr. Burr 2004). This research has not only shown that there is such a migration but also that it happens on several different levels each of which is itself indicative for the value attributed to the linguistic entities by the individual people and / or the respective communities:

1) On a first level dialects appear in the Web exclusively in the form of their history and explanations of their particularities. Such pages are normally written in Italian (cfr. Associazione L.I.S.A. 2004).

\(^1\) That the number of English Web pages is still relatively high is due to the fact, that in a lot of countries, the first language to appear next to the national language on the Web is English.
2) Another way to give dialects a space in the digital world is to put collections of traditional texts like rhymes, popular sayings, proverbs in the respective dialect online together with an Italian translation (cfr. Giuse 2003).

3) On a third level we can position sites which are technologically more advanced, i.e. where the multimedia possibilities offered by the Web are exploited in such a way that you can listen to the sayings and rhymes being spoken in the dialect (see for example the *filastrocche* in Bortoluzzi / Bolzan 2000).

4) On a fourth level we find, instead, sites which do not just aim at the conservation of dialects as testimonies of a lost past or a dying culture such as the ones named so far, but which by exploiting the interactive possibilities offered by the Web make available tools like online dictionaries and forums or chat facilities for the active usage of dialects. One example is *dialettando.com*.

The most complex sites are found, however, at least at the moment, when it comes to minority languages. In fact, communities which traditionally have a strong feeling of cultural and linguistic identity and have long sought to gain recognition as autonomous entities have quickly recognized the possibilities offered by the new technologies for the affirmation of their identity and have thus been present on the Web already for quite some time now. In this case, too, we have to distinguish, however, between several types of communities.

There are first of all communities which exploit the possibilities offered by the new technologies more or less exclusively for the conservation of their cultural heritage. An example is the site of *Union Ladina del Cadore de Médo*, who, by presenting the wealth of historical and cultural information gained by ethnographic, sociological and linguistic research not in Ladinian, the traditional language of the region, but in Italian and by giving a voice exclusively to very old representatives of the Ladinian community, categorizes the minority language as belonging to the past. Another type of communities, on the contrary, with the help of the new technologies aims clearly at promoting its language as a modern means of information and communication. This is the case with *La Patrie dal Friûl*, a monthly journal of cultural and political information where Friulian is actively used. The importance which this community attributes to the present technological development for the affirmation of their own language in the modern world is furthermore shown by the fact that the site not only makes available an interactive Friulian-Italian dictionary but also an interactive multilingual dictionary, where the Italian, Slovene, German, English, Spanish and French correspondences of Friulian words, and vice versa, can be retrieved.

But the process of migration into the digital world does not stop with those communities who have always defended their identity. Instead, also communities who in the past have been much less active on a level which transcends their own locality are now discovering the possibilities of the web and are creating their own sites. Interestingly, newcomers like the Germanophone communities in northern Italy seem less governed by the wish to strengthen just their own identity, rather than by the intention to overcome at the same time the geographical borders which separate the individual communities. The site *Isole linguistiche* is a good example here.

These and many more indications that information and communication technologies do not really serve monolingualism and homogeneity but that they encourage, on the contrary, the recognition of the value of multilingualism and heterogeneity concern, however, only one aspect of the present development, i.e. the presence of more and more national, regional and minority languages, of individual language varieties and local dialects on the Web. Another aspect of the same development would seem to be the creation, affirmation and acceptance of mixed languages or hybrid varieties.
From Multilingualism to Hybridity

Lately, the crossing of language and dialect boarders, the creation of multilingual spaces, the production of new identity patterns and the speaking of several languages at the same time, for which Jürgen Erfurt (2003: 6) has coined the term-multisprech (Multispeech) has more and more engaged the interest of socio- and applied linguists in Germany (cf r. Erfurt 2003). Their studies are, however, mostly confined to spoken discourse in urban migration milieus. Yet, as other research has shown, phenomena of Multispeech are neither limited to urban migration milieus nor to the domains of spoken speech. Instead, it would seem that the new technologies are one of the factors which encourage the overcoming of certain norms and help the (intentional) development of mixed languages even in environments dominated by writing.

3.1 SMS and Multispeech

The phenomenon and its inherent technological determination was first noted, as far as I can see, by Flavia Ursini (in print), when she studied two corpora of SMS written mainly by northern Italians. In these corpora, next to messages written wholly in the dialect of the region as in the following examples:

C1, 89: A MI PAR MI VA BEN, CASA TUA SOITA ORA.
C1, 95: oppo quea ke te gao contao, pi gnente, ovio kendemo. Saeuti. M.

or messages where code-switching between northern Italian varieties and standard Italian takes place as in the following examples:

C1, 29: Dottore, si faccia vivo ogni tanto. So sempre mi che te gho da ciamare.
C1, 34: /.../ è da un bel pezzo che non ti fai sentire. Vien fora da che busa!
C2, 154: /.../ Mi tocca andare fino a M***** col trenino, e poi tornare da li! Ecc! Desperassiu! Ecc! Raffredùr! Ecc! Ciau!
C2, 198: Sper che tut vad ben (poveri bamb.). B***** mi ha detto di dirdi di contattarlo. Ciau. G.

Ursini found, in fact, also quite a number of messages where the writers either mixed their own northern Italian dialect with central- and southern Italian dialects:

C1, 2: Non mi vhenire addire che anche questo sabato te ne vuoi stare acchasa a far finta di studiare!
C1, 7: non è checc’hai tèmpo di passà eddi vvenirmi a prêndere in macchina?
C2, 34: Ciò le foto! Sono ‘ribbili (le mie). Le tue accussi accussi… (invidia…). Pa’ i seminarì te ciamme dopp!
C2, 107: Agge ditt’a ‘o compiuto mio che lo duomilo è ggià arrivate. Chillo m’ha ditto: “Föttiti! A me nun me freghi!” E io, che cce devo penzà? Ah, chiste tecnuologie!

or, and this is what interests most in our context, where a mixed language made up of different languages and dialects appears:

C1, 81: /.../ io non so gnancora perké dovaria studiar /.../ in + dovrei lavarmi i cavei. Fame savuar
That these phenomena have nothing to do with backwardness, uneducated speech or thoughtlessness is shown by the fact that the language used in the other messages written by the same people is the modern middle level Italian or *italiano dell’uso medio* and that in these messages its orthography is generally respected.

Ursini’s conclusion that such a mixing of languages is indicative of a search for a language which is as far removed as possible from the standard language and comes as near as possible to the pole of informality (cfr. Ursini in print: 453) certainly captures an important aspect of this type of technology-mediated writing. I would, however, go further and interpret this type of linguistic practice as an indication of the weakening of the national and standard language paradigm, of the search for a *we-code* which counterbalances the *they-code* represented for most people in Italy by Italian and for a linguistic identity built on the multilingual knowledge of the speakers. Such a hypothesis can, in fact, be supported by research on other hybrid languages and above all on Nouchi.

### 3.2 Nouchi

Nouchi, according to Sabine Kube (2003) was originally created by youngsters in Abidjan (Côte d’Ivoire). In this multi-ethnic and multi-lingual city, French, the official language of the country, functions also as the *lingua franca*. The reason for this situation seems to be that given the very complex linguistic situation, none of the original languages managed to take on this function. This does, however, not mean that French is also attributed an identity-building function. Instead, this function is realised by Nouchi, a hybrid language composed of elements of French, of elements of the African languages present in Côte d’Ivoire and of neologisms invented by the people of Côte d’Ivoire themselves. Nouchi, the usage of which is no longer confined to oral domains but has become the language of music groups, artists and satirical weeklies, seems to constitute, in fact, the *we-code* which opposes the *they-code* represented by either the standard French language which is taught in the schools or by the French *lingua franca*. It is even said that Nouchi could well become the expression of Côte d’Ivoire’s identity altogether in the future (cf. Kube 2003: 131-137).

### 3.3 Europanto

Similar things have been said more than once about Europanto, a mixed language ‘created’, or better, brought to light in 1996 by Diego Marani, an Italian translator who works in Brussels:

Is this a new artificial language, an alternative to Esperanto? Or could it be the language of the future? (Marani 1999)  
L’Europan – Könnte das la lingua majeure for Europa para la nouvelle millennium sein? (Karismo 1999)
Notwithstanding the fact that on The Linguist List Europanto is seen to be closely related with mixed (artificial) languages like Esperanto, Interlingua, Brithenig, Ceqli, Klingon, Lojban, Orcish, Ido, Quenya, Sindarin, Volapük, Jakelimotu (cfr. Linguist List 2004), Europanto cannot really be included in this list, because it was not created for literary or film characters as Brithening, Lojban or Orcish, nor was it created by linguists so to speak from above, like Esperanto, Ido or even Interlingua, which is presented in the following way:

Interlingua es un lingua international facile e de aspecto natural elaborate per linguistas professional como un denominator commun del linguas le plus diffundite in le mundo in le dominios del scientia, cultura, commercio, etc. Un texto in interlingua es immediatemente intelligibile a milliones de personas in tote le mundo, sin necessitate de studio previe (UMI 2004).

Europanto, on the contrary, is as Marani puts it, developing in a natural way from the bottom of the magma of European multilingualism (cfr. Marani 1999). An example of a text in Europanto written by Marani is the following:

G7 INFORMATIOGEZEL PILOTAPROJECTO
Diego Marani 09.04.1997
"Eine globalo kaufpunto por Piccola und Media Entreprisas"

Que would happen if, wenn Du open your computero, finde eine message in esta lingua? No est Englando, no est Germano, no est Espano, no est Franzo, no est keine known lingua aber Du understande! Wat happen zo! Habe your computero eine virus caught? Habe Du sudden BSE gedeveloped? No, Du esse lezendo la neue europese lingua: de Europanto! Europanto ist uno melangio van de meer importantes Europese linguas mit also eine poquito van andere europese linguas, sommige Latinus, sommige old grec. Qui know ten moins zwei europese linguas kan Europanto undergrepen. From nu avanti, Du need keine mas foreignas linguas studie und Du kan mit el entiero mundo communicare danke al Europanto. Du no believe? Ich zal aan you demonstrer brefly describendo en Europanto el pilota projecto "Eine globalo kaufpunto por Piccola und Media Entreprisas". El but del projecto ist de facilitate PME in der electronicommerz. De aczione in tres themas axed esse:
1. Informatio Network por PME
2. PME Exigentias: juridica, institutionale, technica
3. Internazionale testbeds

Europanto, furthermore, is certainly not any longer merely ‘a language of Belgium’ or restricted to the members of the European Institutions as the Ethnologue defines it (ethnologue.com 2004), nor is it used exclusively by Marani for his newspaper columns, novels or translations. In fact, in the meantime quite a few Internet-sites of Europanto have been created which are dedicated either to the collection of texts written in Europanto such as the site Europanto, which claims to possess “Mas than 100 textes” (Anonymous 2004), or to the usage of Europanto, like the forum created in 2001 by Neuropeans where people talk about Europanto in Europanto (neuropeans.com 2004).
4 Relevance for Linguistics and Humanities Computing

Whether Europanto constitutes a similar case to Nouchi in Côte d’Ivoire remains to be seen. The same goes for the question regarding the links which exist between Europanto and (European) identity, as this can only be answered by systematic research into the character of Europanto and the arguments used in online forums in favour of using and spreading it.\(^2\)

What is already now clear, however, is that the whole process from monolingualism to multilingualism and to hybridity, which seems to be an important ingredient of digital culture, raises many questions and opens up new research fields, above all for linguistics. Digital culture is, after all, heavily dominated by writing and by specific tools. With the migration of languages, dialects and hybrid varieties which have not been written before (or at least not according to a unique norm) into domains of digital culture (SMS, the Web etc.) where no academy or other institution exerts control or establishes binding norms, linguistics is presented with a unique possibility to study directly how the ‘oral’ is being transformed into the ‘written’, how norms evolve and spread, which tools are considered to be vital if a language or dialect is to be recognised as being part of the modern world and how these tools are being created.

Mixed languages and their creation and acceptance by speakers, furthermore, question thoroughly those schools of linguistics which in conformance with the idea of a standard or national language have modelled language as a homogeneous entity, the ideal speaker-listener’s competence as inherently monolingual, and have reduced linguistic variety to a common core. It becomes clear, in fact, that such models and theories have hindered linguists from seeing the multilingual nature of human communities, of the linguistic knowledge of the speakers and of their linguistic behaviour. By looking at migration in urban environments, instead, by studying the role new technologies play in freeing people from the control of print and from state norms, by examining the free choices taken by multilingual people and by following up what less main stream linguists like Hymes (1972), Wandruszka (1979) and Coseriu (1988) had to say about the multilingual nature of humans and their communities we will open the way for a quite different conception of languages, of linguistic knowledge and of speaking.

Last but not least, mixed languages and / or hybrid varieties raise a number of questions for Humanities Computing. One question is, for example, how to handle such language texts when we build corpora and above all how to treat the elements of the different languages of which they are composed when we mark up such texts? Such texts are after all not made up just of elements which can clearly be attributed to one or the other language, as in the example of code-switching given below:

Drittens, and not the least, nous le croyons probable, daß die milieux multiculturels avec leur richesse et créativité linguistiques will grow in importance and strength, and thereby achieve the respect necessary für sprachliche Mannigfaltigkeit to become the standard of society in Europe (Jørgensen / Kristiansen 2000: 167),

but also of elements which are being created spontaneously or by taking into account certain characteristic traits of certain languages as in the following quote:

\(^2\) As such research can be carried out systematically only on the basis of a corpus, at the moment of writing a marked-up and TEI conformant corpus composed of the messages which appear in the above presented forum is being created. This corpus will be integrated later with other types of texts written in Europanto.
It was ein ernormer plaisir ihre Europanto™-website zu découvrir (grâce to the New York Times del Web). Saviez-vous que il existiert un phenomenon della same genus an ein niveau Transpacificano? Ici à Los Angeles (place déja at least bilingue), parmi los immigrants, como Sinistas, Koreanos, Japonicos, sans mention Latinoamericanos, une sprache incorporant divers elements vom grammatik y vocabular both from Europa et d'Asien has evoliert. Avec mes freunden, jo speak oft upgemixt Englisch with Chinese wörtorder and etwa Japonais verba, ou bien vice versa. C'en est the same chez youth in Taiwan, who ajoutent auch two dialecta differenta della lingua sinica.

Naturlich sind you Europeans beaucoup mieux organisado, mit wesbites u.s.w. -- and j'applaud ihre accomplissements en esta demesne. But les 'Ricains (though j'admets io non esta) y los Asiaticos might well vous rattraper, sooner or tard.

Finalemente, suggestare un challenge pour les informaticos: ist es possible to développer un spell-checker Europanto™?

En tout K, grazie füe ihre site,
B.R., 25/03/98

And what do mixed languages – which are perhaps much more common than linguistics has realised up to now – mean not only for grammar and dictionary building but for speech recognition as a whole? Can the linguistic knowledge used by the speakers when they produce such mixed language texts be exploited for the processing of natural languages? What does this knowledge tell us about languages and speaking? What sort of tools do we need if we want to study it? And how about a spell-checker and other tools for Europanto and other mixed languages like Nouchi and the many Creoles spoken (and written) in many parts of the world? It will be interesting to see whether this challenge will be taken up in the near future.

Bibliografia


Isole linguistiche <http://www.isolelinguistiche.it/prima.htm> [18.11.2004]


