AARHUS UNIVERSITY

Coversheet

This is the publisher's PDF (Version of Record) of the article.

This is the final published version of the article.

How to cite this publication:

Bakker, P. (2004). Phoneme inventories, language contact, and grammatical complexity: A critique of Trudgill. *Linguistic Typology*, 8(3), pp. 368-375. doi:10.1515/lity.2004.8.3.368

Publication metadata

Title:	Phoneme inventories, language contact, and grammatical complexity: A critique of Trudgill
Author(s):	Peter Bakker
Journal:	Linguistic Typology
DOI/Link:	https://doi.org/10.1515/lity.2004.8.3.368
Document version:	Publisher's PDF (Version of Record)

The final publication is available at www.degruyter.com

General Rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

If the document is published under a Creative Commons license, this applies instead of the general rights.

- Tekin, Talat (1967). Determination of Middle-Turkic long vowels through 'Arud. *Acta Orientalia Hungarica* 20: 151–170.
- (1975). *Ana Türkçede Aslı Uzun Ünlüler*. (Hacettepe Üniversitesi Yayınları, B-15.) Ankara: Hacettepe Üniversitesi.
- (1976). On the origin of primary long vowels in Turkic. *Ural-Altaische Jahrbücher* 48: 231–236.
- (1995). *Türk Dillerinde Birincil Uzun Ünlüler*. (Türk Dilleri Araştırmaları Dizisi, 13.) Ankara: Simurg.

Phoneme inventories, language contact, and grammatical complexity: A critique of Trudgill

by PETER BAKKER

Abstract

There is no support for Trudgill's thesis. There are languages with exceptionally high numbers of consonants that are spoken by large groups, sometimes by millions as a second language. A pilot survey of contact languages (pidgins, creoles, and mixed languages) does not reveal unusual numbers of phonemes. Non-creole languages that have simplified their verbal inflection due to contact and second language learning appear not to decrease phoneme numbers either. Even most mixed languages that arose among fully bilingual groups fail to show exceptionally high numbers of consonants, even though some may display more complicated structures than their component languages. There is no correlation between language contact and phoneme inventories.

Keywords: *bilingualism, creole, language contact, mixed languages, phoneme inventories, pidgin, social structure*

1. The claim

In the target article, Trudgill suggests that language contact can be responsible for small phoneme inventories if adults acquired them in learning a second language (simplification or pidginization), and for large phoneme inventories if there was mostly first language bilingualism. This is a challenging idea, but I doubt that there is enough evidence to support it.

2. Phoneme inventory sizes in the languages of the world

A small pilot survey reveals no correlation between phoneme inventory and the social factors suggested by Trudgill: isolation, contact, and network density. If we rely on the hundreds of phoneme inventories in Maddieson (1984) and select those languages with large phoneme inventories, we encounter a large variety of social situations. I will focus on consonants. The mean number of consonants in the languages of the UPSID sample is 22.8, ranging in all from 6 to 95 (Maddieson 1984: 9). Most languages have between 20 and 37 phonemes (Maddieson 1984: 7).

I arbitrarily chose a subset of phoneme sets in order to select languages with more complex or unusual phoneme inventories. By picking the languages from the UPSID sample with the highest number of fricatives (with the reasonable assumption in mind that such languages will have an overall greater number of consonant phonemes) we get these two: the Northwest Caucasian language Kabardian, with 48 consonants in all, and the Afroasiatic language Margi of Nigeria, with 36 consonants. The languages with the highest numbers of series of stops are the Mexican languages Otomi and Mazahua (both some 42 consonants and around 200,000 speakers), the West African language Igbo (42 consonants, 18 million speakers), and the Khoisan language !Xū (95 consonants, 4,000 speakers). The languages with the highest number of liquids are Irish (44 consonants, 250,000 speakers), Ngarluma (coastal western Australia, 21 consonants, 70 speakers around 1970), and Wangganguru (21 consonants, 8 speakers).¹ These languages as a set do not stand out in that their speech communities are exceptional in the sense of Trudgill's parameters. None of their speech communities is strikingly isolated; network density does not seem extreme for most of them, and some of them are in fact learned as second languages as well. Kabardian has, according to *Ethnologue*, 3 % of second language speakers, which is a considerably higher proportion than for, say, Japanese (27 consonants). Igbo is even "the main trade language" of the region, which implies the existence of many adult learners. Gã in Ghana, which is not a lingua franca, has 29 consonants, 13 fewer than Igbo.

I also decided to check the social situations of languages with individual nasal consonants that are very rare in the UPSID sample, assuming that most of those will have larger than average phoneme inventories. Nine languages have unique nasals such as a "voiced dental-palatal nasal", a "palatalized voiceless velar nasal", or a "pharyngealized voiced velar nasal". These languages include

1. Nick Evans points out that the Iwaidjan languages have the highest number of liquids of Australian languages, namely 8 or 9, out of an inventory of 22 or 24 consonants (see Evans 2000: 98-100). All eight Iwaidjan languages have always been spoken by small groups, today not more than 200 people.

Iai (Loyalty Islands; 36 consonants, 2,200 speakers), Teke (Congo, Bantu; 21 consonants, 200,000 speakers), Irish, Yao (33 consonants, several millions of speakers), Russian (33 consonants, 150 million speakers), and Hindi-Urdu (35–40 consonants, with half a billion first and second language speakers).² Social differences clearly do not seem to play a role here: even though a few island languages and languages of (isolated?) mountain people are indeed included here, we find a very wide range of social organizations of these communities and an extreme range of numbers of speakers. Even two languages that have been second languages for millions of people during many centuries are found here. The number of consonants of these languages is indeed above average (they were selected for that purpose), but the number of isolated or high-contact or tight-network groups among these languages is not remarkably high – on the contrary.

This exploratory survey leaves little or no hope that Trudgill's thesis would be corroborated if one would choose a proper sample and a solid method of classifying the language groups along Trudgill's social parameters. I have absolutely no confidence that the languages with extreme phoneme inventories from that sample would be connected with the special social situations suggested by Trudgill. The observations that the languages with the smallest numbers of phonemes are found in isolated island groups, and languages with large numbers of phonemes in contact situations, are not general, and his cases are therefore likely to be due to chance and areal patterns.

3. Phoneme inventory sizes of contact languages

Trudgill suggests that certain contact languages learned as second languages have smaller phoneme inventories. Creoles and pidgins are the most extreme examples of those. It appears, however, that the phoneme inventories of pidgins and creoles are not significantly smaller than those of non-contact languages.

Many distinctive features such as tone and vowel length tend to get lost in pidginization and creolization. However, I am not aware of any pidgins or creoles that have phoneme inventories significantly smaller than the inventories of the languages spoken in the contact situation. Phonetic simplification does take place, but typically not to a great extent in phoneme inventories. A pidgin/creole like Tok Pisin may have fewer phonemes than English, but this is most likely due to the less extensive inventories in the Papuan languages spoken as first languages.

2. The Australian language Maung (17 consonants, 200 speakers in the 1970s, all bilingual among others in Gunwinggu) is also listed by Maddieson as having a rare nasal consonant, but that appears to be incorrect. (Thanks to Nick Evans for pointing this out.)

Creoles are perhaps the least interesting contact languages in this respect if one follows McWhorter (2001) in his claim that creole languages are simpler than non-contact languages. It is not surprising that this can be claimed, given the genesis and history of these languages out of extreme lack of access to lexifier languages. Although I am not aware of a systematic comparison of phoneme inventories of creoles with those of their contact languages (substrate and lexifier), creole inventories tend to show more of a compromise between the languages spoken by the groups in contact, with a considerable subset of the number of phonemes used in the languages in the contact situation, usually strongly influenced by the first language of the creators of the creoles, or nativizers of the pidgins.

Phoneme inventories of pidgins tend to be close to those of the lexifier languages, or other languages involved in the contacts. Chinook Jargon as spoken by American Indians includes phonemes such as glottalized consonants and voiceless laterals. Its inventory is roughly the same as those found in the neighboring languages of the Northwest coast. Chinese Pidgin English had 25 consonants (of which some were merged with others by some speakers). In pidgins, there is always variation and interference from the first languages of the speakers, and therefore phoneme inventories are not always easy to determine. Mobilian Jargon has 18 consonants, Turku Pidgin Arabic 21, Ghanaian Pidgin English 24, and Fanagalo Pidgin Nguni (Bantu) 34. These numbers approximate those of the lexifiers, or the first languages of their users. Phoneme inventories of pidgins do not seem to be extremely low.

In short, neither in creoles nor in pidgins can one find strikingly large or small phoneme inventories, certainly not in comparison with the other languages of the contact situations. L2 speakers may both increase and decrease the number of phonemes of the lexifiers in pidginization. Pidginization cannot be held responsible for small phoneme inventories, despite Trudgill's claim.

4. Phoneme inventories of mixed languages

More interesting in this respect are the modest-sized sets of new languages that developed in situations of bilingualism called intertwined languages (Bakker & Mous 1994, Bakker 1997: Chapter 7), or bilingual mixtures (Thomason 2001). Trudgill suggests that L1 bilingualism in tight networks would lead to large phoneme inventories.

In a number of cases new ethnic groups have emerged from mixed marriages, where the fully bilingual children learned both languages of the partners and, in certain specific social circumstances, created new languages out of a combination of their two languages. I will briefly discuss both morphosyntax and phoneme inventories of these languages.

Intertwined languages combine the lexicon of one language and the grammatical system of another. Examples of such languages are Media Lengua (Quechua and Spanish), Chindo (Javanese and Malay), Angloromani (English and Romani), Ma'a (Bantu and Cushitic), Ejnu (Turkic and Persian), and some more (see also Thomason (ed.) 1997). In almost all cases, these languages have the same phoneme inventory as the language of the grammatical system, plus or minus one or two phonemes. For example, the mixed language Ma'a has 33 consonants, three more than the non-mixed Bantu language from which the grammatical system is derived. The three extra phonemes are mostly found in what are most likely original roots with original phonemes, with some spread to Bantu words (Van Gijn 2000: 103–105). Mednij Aleut is a mixed language with Russian and Aleut elements, albeit not an intertwined language. Aleut has 26 consonants, Russian has 19 places of articulation and 33 consonants, whereas Mednij Aleut has 29 consonants. The Mednij Aleut phoneme inventory is a kind of compromise between Russian and Aleut: it has the voiced stops of Russian, whereas the Aleut distinction between uvulars and velars is no longer phonemic, and a few marked Aleut phonemes are not used in Mednij Aleut. A newly “discovered” mixed language of Nigeria, spoken in Okrika, combines Igbo verbs with Ijo nouns (Wakama 2000, forthcoming). Igbo has 28 consonants and the mixed language 27. Okrika has a few consonants lacking in Igbo, and a few Igbo consonants are not used in Okrika. In short, there are no extreme differences between the mixed languages and their component languages.

One of the most intriguing examples of a mixed language is Michif, with verbs from Cree and noun phrases from French (Bakker 1997). Michif is spoken by the descendants of male French speakers and female Cree speakers who merged into a new ethnic group, the Métis. The phoneme inventory of Michif is a subject of debate. All researchers seem to agree that the phoneme inventory of Michif is not a compromise between French and Cree, but rather a sum of the inventories of the two languages. One claim is that at least some of the phonemes seemingly shared by the two components (such as /p t k/ and especially the vowels) are in fact distinct phonemes, based on the different phonetic ranges of their allophones. For instance, in the French part /o/ and /u/ are distinct phonemes, but in the Cree part they are allophones. Others would claim that these overlapping phonemes are in fact the same phonemes. In any case, the phoneme inventory is larger than the number of phonemes of either language. Cree has 13 consonants, the Métis French variety that was the source of Michif has 21 consonants, and Michif has 24 consonants if one considers the overlapping part as the same phonemes (French /k/ is the same as Cree /k/). If one counts the phonemes with different ranges of allophones separately, Michif may have as many as 31 distinct consonants – but this remains to be proven, for instance by experimental means. In both counts, the mixed language has more phonemes than the individual source languages.

These examples of mixed languages show that their phoneme inventories do not exceed those of any of the individual languages involved in the contact situation in a significant way, with the possible exception of Michif.

5. Grammatical complexity and language contact

What about grammatical complexity? Would contact lead to morphosyntactic complexity and simplicity? Do languages with contact-induced or socially-dependent inflectional simplification also show a decreased number of phonemes?

Kusters (2003) is a systematic attempt to test whether varieties of a language that have been learned by many people as a second language are morphologically simpler than varieties that have been transmitted from parents to children. The more second language speakers and the less tight the speech community, and the less stable and prestigious the language, the less complex the language is expected to be. By comparing insular Scandinavian Icelandic and Faroese (little contact) with Norwegian (much contact), and different varieties of Arabic, Quechua, and Swahili with each other, Kusters shows that there is a correlation between the simplification of the verbal inflection of the source languages (Old Norse, Classical Arabic, “standard” Swahili, and the variety of Quechua spoken in the community from which the language spread). In Kusters’ method, the original languages and cultures receive the fixed score 10 for complexity. The other communities score between 5.5 and 2 for societal complexity, and their languages between 5 and 1.5 for inflectional complexity, providing convincing evidence for contact-induced simplification in the morphological domain.

If this works for grammatical complexity, one would expect an effect on the phoneme inventories as well, as suggested by Trudgill. Here, however, there is no fixed pattern. Morphologically simplified Norwegian appears to have more consonants than its sisters (Norwegian 22, Faroese 16, and Icelandic 17). Grammatically simplified Ugandan Nubi Arabic creole has 21–30 consonants (the higher figure including loans), and less simplified Egyptian Arabic has 31 consonants disregarding their length. There does not seem to be a correlation between simplification of inflection and the (reduced) size of phoneme inventories.

6. Morphosyntactic complexity in mixed languages

What about morphosyntactic complexity in contact languages with increased phoneme inventories? Morphosyntactically, Michif combines the complex parts of both languages. It uses virtually the complete verbal system of the polysynthetic language Plains Cree, with a template of some 20 morpheme slots. It uses the French nouns, with part of the idiosyncratic (and often still

productive) derivational morphology of French, and intact masculine–feminine and definite–indefinite distinctions of French. Most interestingly, the language combines the masculine–feminine gender system of French (overtly expressed in the determiners and some adjectives) with the animate–inanimate gender system of Cree (overtly expressed in the demonstratives, the stem alternations, and the verbal inflection). Michif is indeed a language of increased phonological and morphosyntactic complexity, compared to both of its component languages.

Another example of a mixed language, or rather a group of three languages, that show increased complexity, can be found on the Reef Islands/Santa Cruz. They display both Austronesian and Papuan features in their lexicon and morphosyntax, and it has been a matter of controversy whether the original affiliation of the languages was Papuan or Austronesian (see Wurm 1986). Wurm (1992: 527) called the phonologies of these languages “fairly complex”, with twenty oral and nasal vowel phonemes and between 17 and 21 consonants. In addition, the verbal morphology probably ranks among the most complex ones of the world, with no fewer than seven prefix slots and 23 suffix slots (Wurm 1992). Nominal classification shows one of the most elaborate systems of the world, too (Wurm 1981). At least 25 possessive classes have been identified, and there may be as many as 50 (Wurm 1986: 519), with three different types of concordance. In fact, the stems of many basic words appear to consist of smaller elements, often similar to noun classifiers in other languages (Wurm 1986).

Michif and Reef Islands/Santa Cruz, however, seem to be exceptions: most mixed languages do not display extreme differences in complexity relative to their component languages.

7. Conclusions

What can we conclude from this pilot survey? First, that there is no doubt that a variety of a language learned by many people as a second language and then transmitted to new generations will indeed lose some of its grammatical complexity and irregularities, or indeed a lot, in more extreme circumstances (pidgins, creoles). But there may not be a simplifying effect on the phoneme inventory. Pidginization and creolization do not significantly decrease phoneme inventories. Second, new languages spoken by descendants of bilinguals who at some point in time combined their two old languages into a single new one normally do not display more complex phonological and morphosyntactic features. Only a few of such contact languages did increase or elaborate the sound systems and the grammatical systems into a new language that is considerably more complex than either of the source languages. Most languages, however, do not show a correlation. Finally, a pilot survey of languages with large con-

sonant inventories appears to reveal a wide variety of social functions for these languages, including having millions of second language learners. Apparently there is no fixed correlation between contact and the size of phoneme inventories. Therefore I remain very skeptical of Trudgill's thesis.

Received: 23 October 2003

Aarhus Universitet

Revised: 26 January 2004

Correspondence address: Institut for Lingvistik, Aarhus Universitet, Nobelparken, Jens Chr. Skous Vej 7, 8000 Århus C, Denmark; e-mail: linpb@hum.au.dk

Acknowledgements: The author would like to acknowledge the Danish Center for Advanced Research in the Humanities for the financial and infrastructural support of my research in 2002–2003, and to thank Nick Evans and Bill McGregor for comments on an earlier version.

References

- Bakker, Peter (1997). "A Language of our Own": *The Genesis of Michif – the Mixed Cree-French Language of the Canadian Métis*. New York: Oxford University Press.
- Evans, Nicholas (2000). Iwaidjan, a very un-Australian language family. *Linguistic Typology* 4: 91–142.
- Kusters, Wouter (2003). *Linguistic Complexity: The Influence of Social Change on Verbal Inflection*. Utrecht: LOT, Leiden: Universiteit Leiden, Centre for Linguistics.
- McWhorter, John (2001). The world's simplest grammars are creole grammars. *Linguistic Typology* 5: 125–166.
- Maddieson, Ian (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press.
- Thomason, Sarah Grey (2001). *Language Contact: An Introduction*. Edinburgh: Edinburgh University Press.
- Thomason, Sarah Grey (ed.) (1997). *Contact Languages: A Wider Perspective*. Amsterdam: Benjamins.
- Van Gijn, Rik (2000). The syllable in mixed languages. M.A. thesis, Universiteit van Amsterdam and Aarhus Universitet.
- Wurm, Stephen A. (1981). The possessive class systems in Äiwo, Reef Islands, Solomon Islands. In *Papers in New Guinea Linguistics No. 21* (Pacific Linguistics, A-61), 181–209. Canberra: Australian National University.
- (1986). Remarks on some language problems in the Santa Cruz Archipelago, Solomon Islands. In Benjamin F. Elson (ed.), *Language in a Global Perspective*, 507–523. Dallas: Summer Institute of Linguistics.
- (1992). Some features of the verb complex in Northern Santa Cruzan, Solomon Islands. In Tom Dutton, Malcolm Ross, & Darrell Tryon (eds.), *The Language Game: Papers in Memory of Donald C. Laycock* (Pacific Linguistics, C-110), 527–551. Canberra: Australian National University.