Language Complexity as an Evolving Variable

Edited by
GEOFFREY SAMPSON, DAVID GIL, AND PETER TRUDGILL
Sociolinguistic typology and complexification

PETER TRUDGILL

1 Background

In previous work in sociolinguistic typology (e.g. Trudgill 1983, 1989, 1992, 1996a, b, 1998, 2002) I have suggested that it might be possible to argue that the distribution of linguistic features over languages may not be just random or areal. Rather, it may be the case that certain types of society, or social structure, tend to produce certain types of language, or linguistic structure. In the context of the present volume, therefore, the suggestion is that certain types of social factor may predispose languages to demonstrate greater or lesser degrees of complexity.

Previous work suggests that amongst the most promising social factors to examine in the typological search for the social correlates of complexification and simplification are the following:

(i) degree of contact v. isolation: the focus of discussion here is on the “critical threshold” (Lenneberg 1967);
(ii) denseness v. looseness of social networks, following especially Milroy and Milroy (1985);
(iii) small v. large community size. Relevant here, amongst others, is the work of André Haudricourt (1961).

2 Contact and the critical threshold

At least one helpful way of looking at the concept of linguistic complexity is to regard it as being synonymous with the difficulty of acquisition of a language,

or a subsystem of a language, for adolescent and adult learners. This approach has received support (e.g. Kusters 2003), but it is of course not the only way of looking at complexity, as other chapters in this volume reveal. In particular it has been queried by Dahl (2004: 282), who suggests that “difficult” and “complex” should not be automatically equated: “what an individual finds difficult obviously depends not only on the complexity of the object of learning but also on the individual’s previous knowledge.” One cannot but agree, but my suggestion is that, other things being equal, some systems are more difficult and therefore complex than others. In a paper written in 1979 (Trudgill 1983: 106) I noted:

It is usual for laymen to claim that some languages are easier to learn than others. Linguists have tended to play down this suggestion, and to point out that it depends on what your point of departure is: Spanish is easier for an English speaker to learn than Chinese, but for a speaker of Thai it might be the other way round. However, I think it is legitimate to suggest that some languages actually are easier for adults to learn, in an absolute sense, than others. If one were given a month in which to learn a language of one’s choice, I think one would select Norwegian rather than Faroese, Spanish rather than Latin, and Sranan rather than English.

Crucially, the preferred languages for rapid learning are all “languages which have undergone more contact” (see also Trudgill 1989). Not only are some languages more complex than others, but also this very often has something to do with the amount of contact a variety has experienced.

The link between language contact and simplification has often been made in the literature. For example, the link has been demonstrated by Kusters (2003), who examines the history of degrees of complexity and simplicity in verbal inflectional morphology in Quechua, Swahili, Arabic, and the Scandinavian languages. His highly detailed quantitative analyses lead him to conclude that “the level of [linguistic] complexity is related to the type of speech community” (2003: 359), in that language varieties with a history of higher contact also tend to demonstrate higher degrees of simplification.

Language contact has this consequence because of pidginization. The most extreme outcome of pidginization is the development of a pidgin language, but this is a very rare occurrence. It is only pidginization at its most extreme, together with a number of other unusual factors, which combine to lead to the development of pidgin and, even more rarely, creole languages. Pidginization can be said to occur whenever adults and post-adolescents learn a new language (Trudgill 1989). This in turn – although other factors such as motivation may be involved – is due to the relative inability of adult humans to learn new languages perfectly. Adult (and adolescent) humans are speakers
who have passed the critical threshold for language acquisition (Lenneberg 1967). The critical threshold is the most fundamental instrumental mechanism involved in pidginization, and it has to play an important part in our understanding of language contact phenomena.

Pidginization consists of three related but distinct processes: reduction, admixture, and simplification. Reduction refers to the fact that, in pidginized form, there is simply less of a language as compared to the form in which it is spoken by native speakers. Admixture refers to interference or transfer, an obvious feature of adult second-language acquisition.

Simplification, as is well known (see Mühlhäusler 1977, 2001), is a rather complex phenomenon, but it refers crucially to three linked processes (Trudgill 1996a):

(i) regularization of irregularities;
(ii) an increase in lexical and morphological transparency;
(iii) loss of redundancy.

All languages contain redundancy, which seems to be necessary for successful communication, especially in less than perfect (i.e., normal) circumstances; but it is probable that some languages have more redundancy than others. Loss of redundancy takes two major forms, corresponding to different forms of redundancy:

(a) the loss of morphological categories (i.e., grammatical categories that are expressed morphologically). Some grammatical categories are more frequent in the world’s languages than others. For example, all the languages of the world have the category person, but of the other categories some are more and others less common. It appears that these categories may be of different statuses or degrees of importance, and that their functions may be less or more clear. And there are also important differences in how the categories may be expressed, the most important difference for our purposes being between morphological (or synthetic) expression, on the one hand, and lexical and syntactic (or analytic) expression, on the other. Sometimes, but not always, loss of the morphological expression of grammatical categories is compensated for by the use of more analytical structures, such as the usage of adpositions instead of case endings. We can speak of loss of redundancy “with or without repair”.

(b) The second type of redundancy loss occurs in the form of reduction in repetition of information. Repetition of information is illustrated for

example in grammatical agreement, where there is more than one signal that, say, a noun phrase is feminine; or in obligatory tense marking, such as when all verbs in a past-tense narrative are marked for past tense. Here, reduction in redundancy will take the form of reduction in the number of repetitions, as in the loss of agreement.

Pidginization involves simplification because high irregularity, low transparency, and high levels of redundancy make for difficulties of learning and remembering for adolescent and adult learner-speakers. Simplification is the direct result of the critical threshold. Small children, as is well known, will learn perfectly any language that they have sufficient exposure to. But the less complexity a language has—technically speaking, the less irregularity, opacity, and redundancy—the easier it will be for post-threshold learners to master.

This does not necessarily mean that simplification occurs only in high contact situations—Dahl writes: “it is unlikely that those changes can all depend on language contact” (2004: 283) rather it is a matter of probabilities. While simplification and complexification may both occur in all languages, it is in the high-contact situations that there is an imbalance between the two, with simplification dominant. This is presumably why Shosted (2006) does not succeed in demonstrating the validity of the “negative correlation hypothesis”, which supposes that if one component of language is simplified then another must be elaborated.

It is also very important not to lose sight of the role of the critical threshold, because high-contact situations of certain sorts can also lead to complexification. Nichols writes (1992: 193): “contact among languages fosters complexity, or, put differently, diversity among neighbouring languages fosters complexity in each of the languages.” This then is borrowed, added complexity. And here of course the contact will have to be of a very particular type, namely stable, long-term, co-territorial, contact situations involving childhood—and therefore pre-threshold and proficient bilingualism (see Trudgill 2008).

3) Community size and social networks

I now suggest that simplification is not an especially challenging phenomenon for students of language change, precisely because it is rather predictable in terms of the less than perfect language acquisition abilities of adolescents and adults. The reverse challenge is much more severe. As Thurston (1994: 603) says, what we need to be able to do is “to explain how complexity arose in languages in the first place”. I argue that the more challenging problem is: in

---

2 I do not discuss phonology here, but phonological transparency (i.e., low levels of allophony) will also be relevant.
which social circumstances do we witness the reverse process, spontaneous (i.e. not additive, long-term contact-induced) complexification? To what extent can we expect to see a greater degree of the spontaneous development of irregularity and redundancy in low-contact varieties? And to what extent can we expect to see a greater degree of the growth (as opposed to the addition, as a result of borrowing) of morphological categories in such varieties?

The answer would appear to be reasonably clear: are we not forced to assume that it is precisely in (now increasingly rare) low-contact situations that we are most likely to find the development of irregularity, redundancy, and morphological categories, because otherwise none of the world’s languages would have them? But why would this be?

This is where the two further sociolinguistic factors cited above, social network structure and community size, come in. Grace (1990: 126) has written:

A language exists in the people who speak it, but people do not live very long, and the language goes on much longer. This continuity is achieved by the recruitment of new speakers, but it is not a perfect continuity. Children (or adults) learning a language learn it from people who already speak it, but these teachers exercise considerably less than total control over the learning process.

A possibility is that in some communities the “teachers” may have more control over individuals than in others. Insights deriving from the pioneering work of James Milroy and Lesley Milroy (e.g. esp. Milroy and Milroy 1985) clearly indicate that small, tightly-knit communities are more able to encourage the preservation of norms, and the continued adherence to norms, from one generation to another, however complex they may be. Coupled with the absence of contact and the effects of the critical threshold, it is therefore to be expected that in such communities complexity will be more readily maintained.

However, it is also not unreasonable to suppose that small community size and tight social networks will also assist the production of complexification. In small isolated communities, change will be slower – witness Icelandic versus Danish – but when it does occur, it is also more likely to be of the complexification type, the other side of the coin of high contact and loose networks leading to rapid change and simplification. I hypothesize that small isolated communities are more able, because of their network structures, to push through, enforce, and sustain linguistic changes which would have a much smaller chance of success in larger, more fluid communities. These would be changes of a relatively complex or unusual type. So not only is there less simplification in low-contact situations, there may also be more complexification.

This suggestion fits in with the observation of Werner (1984: 219) that: "unsere germ. Sprachen zeigen, dass eine kleine, weitgehend isolierte, sozial homogene Sprechgemeinschaft (wie etwa auf Island) am ehesten bereit ist, Akkumulationen an Komprimierung und Komplikation weiterwachsen zu lassen.

And the same kind of idea is also suggested by Braunmüller (1995), who sees morphological opacity as a characteristic of "small languages".

My hypothesis is therefore that it is in small tightly-knit communities that we are most likely to find an increase in irregularity and opacity, and a higher level of redundancy; and that not only are morphological categories most readily lost during (post-threshold) language contact, but also that they are more readily developed in such tightly-knit communities. And there is some suggestion (although the picture is far from clear at the moment: Trudgill forthcoming b) that the major mechanisms involved in the production of these forms of complexity include fast speech processes (Trudgill 2002), grammaticalization (Trudgill 1995), and "exorbitant phonetic developments" (Andersen 1988; Trudgill 1996b).

As far as morphological categories are concerned, creoles provide a wealth of instructive negative evidence for the sorts of morphological-category development we can expect to find in low-contact dense social network languages and dialects. Pre-critical threshold language contact and social upheaval mean that pidgin languages usually lack all morphology: there are no cases, numbers, tenses, aspects, moods, voices, persons, or genders that are morphologically marked. However, the expansion process inherent in creolization—the "repair" of the reduction of pidginization—involved the re-introduction of some of the grammatical categories that have been lost. Creole languages typically have (optional) aspect and tense markers. And they typically have (optional) plural markers. But note that they have no person markers on verbs—they continue the pidgin practice of signalling person by pronouns. There is no morphological case marking—case is marked by word order, even for pronouns. Mood has to be signalled lexically. In particular, creoles also lack features such as switch-reference systems, evidentials, subjunctives, noun incorporation, polysynthesis, trial-, quadral-, and quintal-number, pronom hierarchies, inalienable versus alienable possession markers, and so on. There is no passive voice. And there is still no gender. Creoles lack what Dahl (2004: 2) refers to as "mature linguistic phenomena" of which "we find that the most obvious one is inflectional morphology" (Dahl 2004: 11). It is then no surprise that they lack grammatical gender in particular, since this is a phenomenon which "cuts across lexicon, morphology and syntax", and since "gender, inflectional morphology and syntactic agreement make up an interesting cluster.
of phenomena that all belong to the later stages of maturation processes” (Dahl 2004: 197).

What we are interested in, then, is the type of social situation in which features such as grammatical gender, switch-reference, case marking, noun incorporation, and so on — perhaps changes of the type referred to by Mühlhäusler (2001), following Bailey (1973), as “abnormal” — are engendered. And it will also be relevant to consider when and where repetition of information and irregularity are most likely to develop.

4 Complexification

If it is the case that complexification typically occurs, or occurs more often, in low-contact situations (see Trudgill 2002), then even a high-contact variety like English permits us to explore this type of more opaque linguistic change by doing so comparatively. We can compare the long-established relatively low-contact Traditional Dialects of English with the majority General English varieties, standard and nonstandard. According to Wells (1982), English Traditional Dialects — a term which would appear to be equivalent to labels such as German Mundarten and French patois — are increasingly hard to find, and occur only in parts of England, especially those further away from London, and in southern and eastern Scotland and parts of Northern Ireland; with the addition of Newfoundland and perhaps parts of the Appalachians. The other varieties of English — General English dialects — are in the most part standardized and/or urban and/or colonial varieties which have a considerable history of dialect and/or language contact, and show very many signs of simplification over the centuries.

The relevance of the Traditional Dialects is that we can check to see if the relatively low-contact linguistic situations in which the Traditional Dialects grew up really have led to the development of features which represent the reverse of simplification: expansion in irregularity, increase in opacity, increase in repetition of information, and the growth of morphological categories. And Traditional Dialects are especially important because it is the sociolinguistic situations in which these varieties developed which most closely resemble the situations in which most language varieties have developed over the past tens of thousands of years of human linguistic history.

Here are just a small but maybe suggestive selection of English Traditional Dialect features, taken from dialects in East Anglia and the English West Country, which do not occur in high-contact Standard English, or high-contact nonstandard mainstream varieties, and which illustrate — although this can only be suggestive — the growth of complexity which, I suggest, may tend to depend for its genesis on low-contact linguistic environments.

4.1 Irregularization

Grammatical change may lead to greater irregularity as well as regularity, and it is therefore worth noting that irregularization may be more common in Traditional Dialects than in General varieties. In the Traditional Dialect of the English county of Norfolk, for example, we find a number of irregular preterites which occur in cases where Standard English has regular forms. These include (Forby 1830; Trudgill 2003):

\begin{tabular}{ll}
\textit{hoe} & \textit{hew} \\
\textit{mow} & \textit{mew} \\
\textit{owe} & \textit{ewe} \\
\textit{row} & \textit{rew} \\
\textit{save} & \textit{seft} \\
\textit{sew /sou/} & \textit{sew /sou/} (as of stitches) \\
\textit{show} & \textit{shew} \\
\textit{snow} & \textit{snow} \\
\textit{sow} & \textit{sew /sou/} (as of seed) \\
\textit{thaw (thaw)} & \textit{thaw} \\
\textit{shriek} & \textit{shriek} \\
\textit{wave} & \textit{weft} \\
\textit{wrap} & \textit{wrap} \\
\end{tabular}

A few of these verbs had strong or irregular preterites on their first attested appearance in Old English or Middle English. This is true of \textit{mow} and \textit{row}, while \textit{sew} had both strong and weak forms in Old English. In these cases, then, regularization has taken place in Standard English and in the General English nonstandard dialects, while the Traditional Norfolk Dialect has preserved the original irregularity. The other irregular forms, however, are not historical. \textit{Show}, for example, was a weak verb in Old English, with the preterite \textit{scawed}, as was \textit{thaw}; and the other verbs were also regular from their earliest appearance. The only exception is \textit{owe}, which goes back to Old English \textit{a3an}, with the early OE preterite \textit{ahte}, which later of course gave rise to \textit{ought}, and was superseded as preterite by regular \textit{awede > owed}. So the Norfolk forms \textit{ewe, hew, sew (of sew)}, \textit{shew, snow, thaw, shriek, seft, weft}, \textit{wrap} are all innovations which involved irregularization. The changes to \textit{owe, hew, sew, sew, sew}, \textit{thaw} are obviously the result of analogy with the \textit{blow} class, but the origins of \textit{shriek, seft, weft}, and \textit{wrap} are less clear.

It is often said that some forms of American English, including Standard English, have replaced the regular preterite of \textit{dive}, namely \textit{dived}, with an irregular preterite \textit{dove} — clearly by analogy with the class of irregular verbs.
such as ride/rode. However, this is not a case of the development of irregularity in Standard English. According to the OED, dove is also a form which occurs in dialectal British English, and, more importantly, the verb was a strong verb in Old English, with the past participle dofen.

4.2 Decrease in morphological transparency

Growth in allomorphy represents a clear case of loss of morphological transparency, and thus an increase in complexity. Kusters (2003), for example, argues that the loss of allomorphy in Arabic represents an obvious instance of simplification. I give two examples:

(i) One example from an English Traditional Dialect which, crucially, has not been paralleled in any variety of General English concerns the third person singular neuter pronoun. In nearly all English dialects, including Standard English, this is of course it as both the subjective and the objective form. In the dialect of Norfolk, however, a complication has developed. In stressed position, the subject form of the neuter singular pronoun is not it but that. This is particularly clear in the case of the “weather pronoun”, where no possibility exists that we are dealing with the homophonous demonstrative:

That’s raining.
That’s cold in here.

But there is no doubt that it also operates in genuine pronominalization:

He shew me the cat – that was on the wall.

However, in tag questions, the pronoun takes the form it, as it does also in the objective case:

That’s raining, is it?
That’s cold in here, isn’t it?
The cat? I don’t like it.

This allomorphastic distinction is not mentioned by the posthumously published Forby (1830), who states that he is describing the Norfolk dialect as it was spoken in the period 1780-1800; and he himself uses stressed it in his illustrative examples. And I have found no cases in the fifteenth- and sixteenth-century letters written by the Norfolk Paston family. The distinction may therefore be a relatively recent instance of complexification.

Differences between objective and subjective pronouns are obviously totally unsurprising. The focus here is on the fact that, and on the type of variety where, a differentiation has developed where none existed before.

(ii) In the Traditional Dialect of East Somerset (Ilahainen 1991), a pronominal allomorphy more complex than that found in other dialects has developed, notably in the third person singular masculine. The subject form of the pronoun is the expected he:

He’s older than what I be

The object form is ’n /an / ë/ which generally functions just like him in other dialects:

I looked up to un and said “What’s say?”

However, there is also an additional form of the subject pronoun which Ilahainen refers to as a question clitic, and which occurs obligatorily in tag questions and other inversions. It is not immediately obvious that this form derives from he, him, or ’n. Ilahainen writes it as her or er and phonologically it is /æc - ë/:

He do live in Latcham, don’t er?

This type of pronominal allomorphy is of course not unknown at all in the world’s languages. The point is that, where it has developed out of an earlier more simple system in English, it has been in a Traditional Dialect that this has occurred.

4.3 Increase in redundancy (a): the growth of new morphological categories

(i) A few Traditional Dialects in a small area of the southwest of England saw the development of a new and fascinating marking of the difference between transitive and intransitive infinitives. Intransitive infinitives (and objectless transitives) in these dialects were marked by the word-final morpheme –y, while transitive infinitives were unmarked. So in Dorset we find (Gachelin 1991):

Can you zew up thesea zem? “Can you sew up this seam?”

versus:

There idden many can sheary nown “There aren’t many who can shear now”.

This is unparalleled anywhere else in the English-speaking world, and is quite possibly unparalleled anywhere else at all. My enquiry on the LingTyp list asking for examples of other languages which have morphological marking for intransitive but not transitive infinitives received three answers, none of them producing a precise parallel. The Australian Aboriginal language Warrgamay has two sets of inflections – those that can attach to intransitive verbs and those that cannot attach to intransitive verbs – but this applies to all verb forms and not just infinitives. A similar situation applies in Ulwa, a Misumalpan language of Nicaragua. The closest we can come to it is in the
Anywa language, a Nilo-Saharan language of Ethiopia and Sudan, where special marking does occur on intransitive infinitives, but only on those which are derived from transitives, which are then marked relative to their transitive counterparts: “weave [trans.]” versus “to do weaving” (Reh 1996: 187).

(ii) A number of Traditional Dialects in the southwest of England developed an interesting phenomenon described by Ihalainen (1976) in which there is a category distinction between habitual verb forms such as:

\[
\begin{align*}
\text{I do go there every day} \\
\text{I did go there every day}
\end{align*}
\]

versus:

\[
\begin{align*}
\text{I goes tomorrow} \\
\text{I went last week}
\end{align*}
\]

This of course is a distinction between two categories which is common enough in languages of the world. In English, however, it represents an innovation, and one which is unknown in any of the General English varieties.

(iii) Also in the English southwest we see the development of the expression of a pronominal category difference between count and mass nouns, such that inanimate count nouns are pronominalized with he but mass nouns with it:

\[
\begin{align*}
\text{Pass the loaf – he’s over there}
\end{align*}
\]

versus:

\[
\begin{align*}
\text{I likes this bread – it be very tasty}
\end{align*}
\]

This too is unparalleled elsewhere in the English-speaking world, except in Newfoundland, where it is known to have derived from the English southwest (Paddock 1991) – it is once again unknown in any variety of General English.

4.4 Increase in redundancy (b): introduction of repetition of information

(i) One example of introduction of repetition of information can be found in the Traditional Dialect of Norfolk, where double tense marking or “past-tense infinitives” can be found. For example:

\[
\begin{align*}
\text{Have the fox left?} \\
\text{No that ain’t, do Bailey would’ve let them went.} \\
\text{“No it hasn’t, or Bailey would’ve let them [the hounds] go” (Trudgill 1995)}
\end{align*}
\]

(i) These same Norfolk dialects also have an overtly expressed second person pronoun in imperatives:

\[
\begin{align*}
\text{Come you out! “Come on [sing.]”} \\
\text{Come you on together! “Come on [plur.]”} \\
\text{Shut you up!} \\
\text{Be you quiet!}
\end{align*}
\]

Yet again, this is not unusual, but it is an innovation in English, and one which has not occurred in General English varieties.

§ Conclusion

I have argued that societal type influences language type. I have argued that language contact involving widespread adult language learning leads to an increase in simplification including loss of morphological categories. (Long-term coterminous child language learning contact, on the other hand, leads to added complexity.) And I have argued, more hypothetically, that small community size and isolation may promote the spontaneous growth of morphological categories; and that they may also promote the growth of irregularity, redundancy, and low transparency. I have cited a small number of examples which might point in that direction, but of course very many more such examples would be needed before we could feel at all confident about the legitimacy of the hypothesis.

If widespread adult-only language contact is a mainly post-neolithic and indeed a mainly modern phenomenon associated with the last 2,000 years, and if the development of large, fluid communities is also a post-neolithic and indeed a mainly modern phenomenon, then according to this thesis the dominant standard modern languages in the world today are likely to be seriously atypical of how languages have been for nearly all of human history. We have become so familiar with simplification in linguistic change – in Germanic, Romance, Semitic – that it has been tempting to regard it as normal – as a diachronic universal. Maybe, however, it is complexification that is more normal. Or rather, we should say, was more normal; some languages, it seems, are certainly more complex than others, but the current diachronic trend is in the direction of an increasingly higher proportion of languages which are increasingly less complex. A number of the English Traditional Dialect features illustrating complexification listed above have already disappeared, or are in the process of disappearing (e.g. Gachelin 1991).