Complexity in linguistic theory, language learning and language change

Wouter Kusters
Meertens Instituut

In this paper I discuss how the notion of complexity can be defined and operationalized to serve as a concept in linguistic research domains like typology, historical linguistics and language contact and acquisition studies. Elaborating on earlier work (Kusters 2003) I argue that a relative notion of complexity is to be preferred over an absolute one. With such a substantial notion, I show that possible objections raised against the concept of complexity are not valid. I work this further out for complexity in verbal inflectional morphology. Finally I demonstrate some intricacies of complexity with examples from variation and change in Quechua varieties.

1. Introduction

In this paper I discuss the notion of complexity and give an example of how it can be used in historical and contact linguistics.

The idea that all languages are complex, though some languages more complex than others, meets with quite some resistance. As in this allusion to Orwell's Animal Farm, many linguists and non-linguists alike consider the property of "being complex" as equal to "having a high value (in a cognitive, social or cultural sense)". Indeed, music with complex harmonies or literary texts with complex interpretations are often appreciated more than their opposites. Such aesthetic considerations supported by chauvinistic reasons led many 19th century linguists to suppose that languages like Latin, Sanskrit and German would be more complex, and that more complex languages would relate to more complex, higher cultures. A reaction to such Eurocentric views came with early American anthropology and linguistics, as exemplified by Sapir's (1921: 219) frequently quoted passage:

> Both simple and complex types of language of an indefinite number of varieties may be found spoken at any desired level of cultural advance. When it comes to linguistic form, Plato walks with the Macedonian swineherd, Confucius with the head-hunting savage of Assam.

For a long period of time, notions of complexity would no longer be used in discussions of the cultural embedding of language.

However, after the expansion and specialization of linguistics into subbranches in the second half of the 20th century, the term complexity is no longer avoided in all contexts. For example, in second language research, Klein and Perdue (1997) wrote on the differences between basic, communicatively barely efficient languages of second language learners and complete languages, as spoken by native speakers and used the term complexity in that context. In dialectology complexity and simplification are also common terms (cf. Andersen 1988; Trudgill 1986), while in contact linguistics (e.g., Mühlhäusler 1974; Winford 1998) simplification is used as a notion without immediate association to evaluative judgements as long as it does not take a too prominent position. However, when discussing languages as wholes, and when considering one language as more or less complex than another, many linguists still think that this idea stems from ill-informed laymen at its best, or is based on nationalistic or eurocentric assumptions at its worst. Nevertheless, in the last years various articles have been published, in which differences of overall complexity between languages are discussed (cf. e.g., Braunmüller 1984, 1990; Dahl 2004; McWhorter in Linguistic Typology 2001, and the comment articles therein; Thurston 1982, 1987, 1992; Trudgill 1992, 1996, 1997, 2001).

In this paper I argue for a feasible definition of complexity, that does justice both (1) to the subbranches of linguistics as described above, where complexity is already used as a term, (2) to laymen who hold intuitive ideas about simple and complex languages, as well as (3) to linguistic theory.

I first spend some words on the reasons why the notion of complexity is, or at least, has been so controversial (Section 2). Next I discuss the difference between "absolute" and "relative" notions of complexity, and argue for the latter (Section 3). In Section 4, I unfold my definition of complexity and comment on some objections against it. In Section 5, I further develop the notion of "complexity" in order to give an example of how it can be applied to variation in Quechua verb morphology in Section 6.

2. Absolute and relative complexity

In the modern literature on linguistic complexity there are two positions. On the one hand complexity is used as a theory-internal concept, or linguistic tool, that refers only indirectly, by way of the theory, to language reality. On the other hand, complexity is defined as an empirical phenomenon, not part of, but to be explained by a theory. This difference, and the confusion it may lead to, can be compared for instance with the difference between a theory that *uses* the concept of emotion to explain – possibly non-emotional – behaviour, and a physiological theory *about* emotions as empirical phenomena, which explains them possibly in non-emotional terms. This distinction should be carefully kept in mind in next description and comparison of the two positions.

The oppositional terms of absolute and relative complexity are coined by Miestamo (2006), who uses the term "absolute complexity" as *not* related to the experiences of a particular kind of language user. Instead, absolute complexity is considered to be an

aspect of a language as an autonomous entity. Since we can discuss such an aspect of an abstracted state of language fruitfully only within some kind of conceptual framework, the notion of absolute complexity turns out to be determined by descriptions and theories of language(s). In this sense complexity and related notions like markedness or naturalness have been widely used in various frameworks (cf. Dahl 2004).

For instance, in level-based phonology we may conjecture that a phonological description that needs an extra level is more complex than one that is describable with only one level. Then we may distinguish complex and simple phonologies in terms of this particular framework, depending on the amount of extra machinery or levels needed. However, whether the actual sounds and structures of the languages in question are also experienced as more complex in terms of acquisition speed, production ease or any other measure remains a separate question, as also Miestamo concedes (2006: 4).

So far so good, however, the distinction between the absolute and relative approach is less clear-cut than it seems. Consider another example: in generative grammar it is assumed that children acquire grammars by filling in parameters. Now, it has been suggested (cf. Hyams 1987) that each parameter may have a default, or unmarked value, and a more marked value, that may be called more complex. Some claim that SVO order is less marked than any other order (cf. Gadelii 1997), because SVO order would be the result of an unmarked, default, or *not complex* parameter setting. What particular parameter this would be in this case is not at stake here. In this example, again, a language as autonomous entity may be characterized as having a complex or simple grammar in terms of the theory about the language. Again, whether the production of SVO orders would be faster or otherwise easier is irrelevant.

In this last example, however, we can no longer keep away from the problem which involves the *status of a grammar*. When a grammar is considered to be only a tool of a linguist to provide an appealing, short and clear description of a language, then indeed the complexity of the grammar has nothing to do with possible complexities in language use. This position with respect to language and grammar is rather awkward, however. Since the 1960s most linguists agree – if not in practice then at least in principle – that grammars model knowledge architecture, cognitive make-up or processing features of language users. In generative grammar it has been stated most explicitly that the quest for universals is *not* a quest for statistic generalizations in the void, but a search for the real innate language acquisition device. Language universals are not only universals in the mind of the linguist, but also found in language users' minds. Although less explicitly, in other frameworks too it is common knowledge that what is described is some aspect of language users.

If so, all parts of our linguistic theories should refer – in the end – to aspects of language users. Still remaining within the framework of generative grammar and the example above, we must assume that the difference between an "unmarked" parameter setting in SVO against a marked parameter setting in e.g., OVS, is reflected in the child's acquisition process, although this may be in a quite indirect way. Nevertheless, when it is claimed that a linguistic theory models aspects of the cognitive or, more

specifically, linguistic make-up of a speaker's mind, the notions within the theory should be reflected in the data as well.

Proponents of the absolutist approach do not necessarily use the term "complexity" in this way. When discussing language features from a typological perspective the relation between linguistic (i.e., typological) theory and actual language use may indeed be indirect or apparently irrelevant. Regularities and patterns may be found on a highly abstract level. For instance, SVO languages are more likely to also have the order Noun-Genitive than Genitive-Noun. It is not immediately clear how such tendencies would be properties of the individual language user. It may even be claimed that these are "emergent properties" that can not be reduced to the domain they originate from. When discussing language on such an abstract level it may be tempting to suggest that theoretical (typological) concepts are independent from actual language use or speakers' minds. If so, complexity may refer, for instance, to the number of rules that is needed for a particular description. We could call this approach *formal absolutist*, and when using it consequently it is a tenable position. For instance, in terms of mathematical theory some grammars can be called more complex, in the sense of needing more complex rules. A context-sensitive grammar is in this sense more complex than a context-free grammar, and a rule that says: "yes/no questions are related to their affirmative counterpart by a complete reversal of all syllables" would be simpler than a rule that relates these two by assuming complex syntactic structures. By defining complexity in this manner we have a term at our disposal by which we can handle and compare various formal theories. However, one cannot have it both ways. When defined in a *formal absolutist* way, it is only a sheer coincidence if complex grammars are also experienced by language users as complex. The formal absolutist simple rule for making yes/no-questions does not exist in human language, and may be considered to be too complex in a more substantial (relativist) way. Defenders of the absolutist formal approach should have no problems with that fact.

If so, the meta-level of the description is clearly separated from the object level. With the term complexity, however, it is felt as somehow a draw-back of the theory, if complexity is *only* a theory-internal notion. This can be illustrated with an example of two theories explaining the same phenomenon. In Quechua varieties there is variation in the inflectional verbal conjugation (see also Section 6). Quechua varieties vary both in the categories expressed, as well as in the way they are expressed. Now, the verbal inflection of Bolivian Quechua – genetically closely related to Cuzco Quechua – has been discussed in two morphological theories. Van de Kerke (1996) discusses Bolivian Quechua verbs within the theory of Distributed Morphology, while Lakämper and Wunderlich (1998) discuss it within their constraint-based minimalist morphology. Leaving out the numerous intricacies of Quechua inflection and the theories involved, the following examples are relevant:

Bolivian Quechua:

(1) *maylla-wa-rqa-yku*
wash-1-PAST-1.PL
'You(sing/plur)/he/they washed us.'

Cuzco Quechua:

(2) *maylla-wa-rqa-nki-ku*
wash-1-PAST-2-PL
'You (sing/plur) washed us.'

(3) *maylla-wa-rqa-n-ku*
wash-1-PAST-3-PL
'He/they washed me/us.'

Now we have two theories according to which we can decide what morphology is simple. According to a theory that focuses on the way syntactic features and morphological spell-out are related, the first example would be more complex. This is argued by Van de Kerke (1996), when he discusses the Bolivian data in the framework of Distributed Morphology. The problem with the Bolivian Quechua data would be, then, that the feature of [+ 1st person] is expressed twice, both in the *wa*-morpheme, as well as in the pluralizer *yku*. From the perspective of Distributed Morphology the Cuzco data are more straightforwardly explained. Subject and object affixes have their own slots, and are spelled out once. Van de Kerke (1996: 130) says:

> We have seen that Cuzco Quechua has a very transparent Agr/Tense system. . . it complies with the Mirror Principle in realising a good match between the order of morphemes and the morpho-syntactic categories expressed. However, this ideal transparency has become opaque by a minor reinterpretation of first and third plural marking in Bolivian Quechua, which not only led to a great number of underspecified and doubly specified surface realizations, but even to the realization of subject markers as pluralizers in the case of *-yku* as in *-wa-yku* [unspecified sub- > 1Plob].

Other morphological theories are less focused on syntactic-morphological (mis) matches, and more on the lexicon-morphology-interface. Lakämper and Wunderlich (1998) use such a "pre-syntactic lexicalist theory". These theories load up the lexical material with features that are dealt with in syntax in other approaches. When morphologies are evaluated in such theories, problems – or, "complexities" – as described above count less. Instead, the number of allomorphs, homonyms and fused morphemes become relevant. From that perspective Lakämper and Wunderlich (1998: 147) draw the following conclusion about Quechua complexity:

> . . . Only when we come to the post-Cuzco stages (like Potosí [this is Bolivian Quechua, WK] or Santiago del Estero) . . . has this kind of deficiency been overcome and an affix-oriented system is produced. However, there can only be small changes, and the new system has to work with the affix material inherited from the former stages. As we have seen, the potentially symmetric system that emerges in the most recent stages of Quechua is not ideal either.

Lakämper and Wunderlich acknowledge that Bolivian Quechua is not ideally simple. However, they assess the complexities of Bolivian Quechua in a quite different way. The double spell-out of 1st person in 1) is not a problem for them, as it is for Van de Kerke, while the number of affixes is more of an indicator to them.

The point of this discussion is that on the basis of the same data it is possible to arrive at different conclusions about what forms are as complex. And in that sense, "absolute" complexity is not so absolute as the name suggests. Absolute complexity, in fact, depends on the theory and perspective from which we look at a particular language or language phenomenon.[1]

The absolutist may either accept the theoretical diversity, and select a theory. Then complexity remains firmly part of a particular linguistic theory, claims no link with empirical complexities in itself, and becomes a formal absolutist notion. Otherwise he may try to make an informed choice in what theory describes the data best. That is, he may confront the two claims on what counts as complexity, with what complexity is *in the empirical data*. In the latter case, the absolutist becomes in fact a relativist, since he makes his theoretical description and notion of complexity depend on some kind of *real* complexity.

3. Definition of complexity

3.1 Points of relativity

I will now turn to the relativist position and how it can be further elaborated. For the relativist, the main question is, relative to whom or what? The relativist position always assumes a perspective or a point of view from where a certain phenomenon is evaluated as complex or simple. Actually, the absolutist position can be considered as a relativist position in disguise. The absolutist position may be characterized as relativist since complexity is evaluated from the point of view of a particular linguistic theory. As discussed above, complications in the absolutist position arise when we acknowledge that a linguistic theory is a model of a language user.

What are the possibilities for a relativist definition? We may choose various kinds of language users and various aspects of language. We could examine what kinds of sounds in Russian are complex for adult Chinese classroom learners. We could examine what morphology of Latin is most complex to handle for readers of Cicero. Or we might consider what kinds of word orders are difficult for speakers of SVO languages when confronted with a VSO language.

---

1. Special cases are languages about which most or perhaps even all linguistic theories come to the same conclusions with respect to complexity. As noted by a reviewer, Ndyuka will be estimated as less complex than Kabardian under any linguistic analysis. Proponents of complexity theory may adduce such extreme examples to defend their idea. However, the displacement of the notion of complexity from any theoretical framework leaves them defenseless on the broad middle-ground of languages, where the measure of complexity is dependent on more fine-grained theoretical decisions and the perspectives of the language users chosen, which leads necessarily to a notion of "relative" instead of "absolute" complexity.

My conception of complexity elaborates on the naive language user's intuitions when he wonders about what languages are complex or simple. This naive language user conceives other languages as non-native languages, and he has an outsider perspective on the speech community of the language in question. Then we still have numerous conceptions of complexity: a native speaker of Chinese will count other phenomena as complex in German, Vietnamese or Swahili, in comparison with a speaker of Turkish or Warlpiri. In order to construct a general notion of complexity, I have defined an ideal language user type, namely a "generalized outsider". This person speaks a first language, and is not familiar with the second language in question, nor with the customs and background knowledge of the speech community. He or she is primarily interested in using language for communicative purposes in the restricted sense of the word, and not in all kinds of verbal play, ritual exchanges, and poetic uses. We model this person as a *generalized* outsider, in order to prevent either facilitating or hampering influences of the first language on acquiring the second language.[2]

Now, I define complexity as the amount of effort a generalized outsider has to make to become acquainted with the language in question. I distinguish three dimensions of language processing where outsider complexity plays a role.

A generalized outsider learns the language in question at a later age, and is not a native speaker. Therefore, phenomena that are relatively difficult for a second language learner in comparison with a first language learner are the most complex. Phenomena that are easy to acquire for a second language learner but difficult for a first language learner are the least complex. The notion "generalized" prevents positive and negative interferences of an accidental first language to cloud our account of complexity.

A generalized outsider does not have much shared (linguistic and non-linguistic) background knowledge with other members of the speech community. She will therefore be helped with a language that is relatively easy to perceive and understand. In contrast, production difficulties of a language are less hindering for an outsider, because she may adapt the difficult production phenomena to some kind of imperfect second language form, while she can of course not force other native speakers to adapt. A language in which perceptual processing is relatively easy in comparison with production is therefore less complex under my definition.

---

2. As several commentators remarked, the notion of a *generalized outsider* may be just as problematic an abstraction as Chomsky's ideal speaker-listener in a homogeneous speech community. However, this notion seems to me to be useful; perhaps it is not valid as an abstraction or generalization over actual concrete cases of *outsiders*, but it is still useful as a kind of most extreme type, or possibility. Indeed, full objectivity as a "view from nowhere" is impossible in science (cf. Nagel 1986), and a complete outsiders' perspective is even more improbable in actual speech communities. Nevertheless, such concepts may play at least an heuristic role in theory building.

A generalized outsider is primarily interested in clear transmission of information, and less interested in learning a language for all kinds of symbolic meanings, like expressing personal and group identity and aesthetic feelings. Phenomena that are not or less functional in this respect are therefore more complex under my definition.

When pondering on what parts of language to select for a more detailed examination, several considerations play a role (cf. Kusters, 2003: 12–21), and I have chosen the domain of verbal inflectional morphology. From a practical point of view complexity in inflection is easily assessed cross-linguistically. In most descriptive grammars information about verbal inflection is given. Secondly, it is more independent from pragmatics, discourse, register, or inter-speaker variation than other linguistic domains, and is rather uniform across various situations. Moreover, it occurs in most sentences, and must be learned and used by all kinds of language users. Nevertheless, it varies considerably *between* languages: some languages have extensive inflectional systems, while others have no inflection at all.

In comparison with derivational morphology, inflectional morphology is less involved with questions of lexical semantics or indeterminacy with respect to the predictability of meaning. In addition, inflectional morphology concerns morphological structure, which is, with its idiosyncratic allomorphies, exceptions, and communicatively superfluous distinctions, less susceptible to trade-off effects than syntax or phonology. Inflection is also quite easy to define independently from theoretical considerations. Between morphological theories there is rather much agreement at least about what counts as inflection. Finally, verbal inflection stands out above other kinds of inflection, since it is most wide-spread both within and across languages.

3.2 Objections against outsider complexity

Now that we have a more detailed definition of complexity, we can answer some objections that have been raised against the notion that there could be differences in complexity.

Many of the arguments are a priori, and boil down to the claim that all languages are necessarily equally complex, because complexities in one domain of language are balanced by simplicity in another domain. For instance, Aitchison (1991: 214) writes:

> A language which is simple and regular in one respect is likely to be complex and confusing in others. There seems to be a trading relationship between the different parts of the grammar which we do not fully understand.

What is usually lacking is a specification of what complexity means. When it is "outsider complexity", most of the objections against the relevance of the concept evaporate, as I will show, while other objections were already incoherent by themselves. Since these arguments are met again and again, lets spend some words on following these arguments to their (absurd) conclusions.

Focusing on Aitchison's trading relationship lets try to defend it by supposing that it is due to limitations on mental capacities; the mind/ brain can process and contain only a limited amount of complexity. An increase in complexity in one component, then, necessarily leads to a decrease elsewhere. This argument presupposes that the average speaker already uses the maximal amount of the "complexity space". Accepting this for the sake of the argument, other problems remain. When assuming that mono-linguals use the complete "complexity space", the acquisition of foreign languages becomes difficult to explain. It would entail either that the acquisition of a second language has the effect that the native language becomes less complex, or that the acquisition of a second language occupies a quite different "complexity space".

Still playing the devil's advocate, and assuming that for each language there is a separate and restricted mental space, problems remain. That is, when we conceive language competence as consisting of lexical and grammatical knowledge, it is hard to imagine how lexical knowledge would be always equally complex. For example, knowing approximately 66.000 words, as Shakespeare had allegedly at his disposal (cf. Efron and Thisted 1976), would in some way have to be equally complex as knowledge of a very rudimentary lexicon restricted to, say, small-village pre-modern agrarian life.[3] One might be inclined to argue that indeed these two kinds of knowledge are equally complex, because with both kinds of lexicons it is possible to *function* in certain niches in society. However, in that case we reach the bottom-line of argumentation, since this entails that all languages that exist must be equally complex just because they exist, and therefore "function". According to my definition of complexity however, a large lexicon, ceteris paribus, is probably more complex than a smaller lexicon, since an outsider has less problems in acquiring and using a smaller lexicon, cf. however, the discussion by Fenk-Oczlon & Fenk (this volume), Juvonen (this volume) and Riddle (this volume).

Another argument is that the trade-off lies in language use somehow: when in one domain of a language complexity disappears, somewhere else new complexities would have to surface. Languages would always remain on the same level of complexity. That is, a simple language cannot exist, because different language components have counteracting "preferences" for simplicity. A structure that is easily produced would be complex to perceive and vice versa. The force of such arguments depends partly on the definition of complexity. For complexity, as I define it, perceptual simplicity outweighs articulatory smoothness. Therefore, when the trade-off is argued to lie between different kinds of language use/users, the objection does not hold, since in my definition complexity refers to only one kind of language user: the outsider.

---

3. It might be argued that pre-modern farmers do not have necessarily small lexicons. However, that would be missing the point: an advocate of equi-complexity should give no empirical evidence of equally large lexicons, but must give a priori arguments why differences in lexicon size are impossible.

A valid argument against the concept of complexity needs to show that there is *always* a trade-off between different domains. However, this seems not to be the case. For instance, the replacement of a vast array of lexically conditioned plural markers as in Classical Arabic by a system with transparent optional plural marking as in KiNubi Arabic does not seem to involve any trade-off. Whether Classical Arabic plural marking is indeed more complex (for an outsider) than KiNubi Arabic plural marking must be shown in empirical tests, but this possibility cannot be excluded beforehand. With respect to a trade-off between phonological components, there are indications (cf. Maddieson 1984: 23) that even the opposite may be the case.

Another flaw in a priori arguments based on a mysterious trade-off is that they say nothing about complications, only about simplifications. It is less plausible that a complication in one part, for instance, the borrowing of words with previously unknown phonemes, necessarily results in a simplification elsewhere. When the trade-off would only apply to simplifications, differences in complexity cannot be excluded by the trade-off argument.

It may be maintained that there must be a trade-off *somewhere* in the language, even if we do not know where. The simplification of plural marking in the example above, may lead to more difficulties in the interpretation of plural forms. The argument in its general form runs as follows: in some languages complexities may be found on a tangible linguistic level. In other languages complexities emerge in an undefined part of the "pragmatic component". In such arguments the trade-off function extends from core language competence to wider communicative and cultural competence.

Such equi-complexity seems not very likely, as long as it is not shown how mediations in complexity between dispersed linguistic domains could take place. When the trade-off would be supposed to lie between a limited number of semantically or functionally related domains, the claim would be more reasonable (cf. Sinnemäki's article in this volume). Although the argument of *a priori* equi-complexity cannot be excluded, it is not falsifiable: for every change in complexity it can be argued that there is another component, in some hitherto unknown domain of language structure, pragmatics, or culture, where the amount of complexity would be leveled out.

4. Methodology

4.1 Three aspects and three dimensions

Now we have discussed and defined complexity, how can we further operationalize it in order to study it in the wild?

First of all we must examine what aspects of inflectional morphology are in fact experienced as troublesome by this "generalized outsider". For that purpose I have distinguished three aspects of the language outsider: the outsider as a second language learner, as a hearer, and as someone mainly interested in the communicative instead

of the more symbolic aspects of language use. I also distinguish three aspects in inflectional morphology (cf. Kusters 2003):

1. The number of categories expressible in the verb; I call this *Economy*. The less inflectional categories and category combinations a language allows, the more economical the language is. For instance, Ecuadorian Quechua is more economical than Cuzco Quechua, since it does not express 2nd person object agreement nor 1st person plural object agreement.
2. The transparency of the expressions of the categories, which I call *Transparency*. The more deviations there are from a most transparent structure, in terms of allomorphy, fusion, fission (that is, the opposite of fusion: one meaning, expressed in two morphemes), and homonymy, the less Transparency. For instance, Argentinean Quechua is more transparent than Bolivian Quechua since it expresses 2nd person object agreement more often with a uniform affix, instead of a fused affix.
3. The consistency within the order of the expressed elements; I name this *Isomorphy*. The more the order in the morphological domain is computable and motivated by the order in the semantic or syntactic domain, the more isomorphic the morphology is. Ecuadorian Quechua is more Isomorphic than Cuzco Quechua, since it has a univocal affix order, that mirrors a semantic-syntactic order elsewhere, while the description of Cuzco Quechua morpheme order needs some extra stipulations.

When we want to find out what outsider complexity is with respect to inflectional morphology, we must examine empirical data on the difficulties of each of the three outsider aspects (second language learning, perception, communicative use) crossed with each of the three morphological dimensions, which yields nine classes. Here the absolute and relative approaches fundamentally differ. In the relative approach, it is in the end an empirical question whether an "outsider" has more problems with Ecuadorian Quechua, Bolivian Quechua, Turkish or Russian inflection. Nevertheless, in practice many of the hypotheses of the absolutist approach about the difficulties of irregularities correspond to empirical findings. This is, however, not so by definition. In the relative approach we can – and in fact we do – distinguish between the problems of first language learners, and compare these with the problems of second language learners. The relative approach allows us to say that, for instance, little Economy is, quite unexpectedly, found to be easy for L1 learners, while it is difficult for L2 learners and "outsiders" (cf. Kusters, 2003: 48ff.).

Empirical investigations into the difficulties of the nine sorts of language processing, are unfortunately not as abundant as we would like. There is, for instance, hardly any research in the perceptional difficulties of an isomorphic uniform order in comparison with variant morpheme ordering. When we try to gather and use the empirical data in these domains as much as possible we come to the following conclusions about what counts as difficult. For the research and data on which this table is based I must refer to my dissertation (Kusters 2003: 45–63).

Table 1. Schematized representation of preferences for inflectional phenomena in various processing dimensions (+ = preference, 0 = neutral and -, --, and --- = degrees of difficulty)

	Speaker	**L1 learner**	**Symbolic use**	**Hearer**	**L2 learner**
Redundant agreement	--	+	0	+	--
Non-redundant agreement	--	+	+	-	--
Aspect/Tense/Mood	-	+	+	0	-
Voice	0	+	+	+	-
Morphological allomorphy	-	--	+	-	---
Accidental homonymy	0	--	0	--	---
Fission	-	-	+	+	--
Fusion	+	0	0	0	-
Phonological allomorphy	+	-	+	0	--
Structural homonymy	+	+	0	0	+
Isomorphy	+	0	0	+	+
Marked affix order	0	0	0	0	0
Inconsistent affix order	-	0	0	-	-

4.2 From the individual to the social

The next step is to examine what kind of languages are simpler or more complex. A straightforward hypothesis is that a language (as a social fact, that is, as "E-language") becomes simpler the more it has been confronted with demands of generalized outsiders. Assuming that the actual state of a language (E-language) is the product of all language users, who produce it on the basis of how they got acquainted with the language in question, it is not too far-fetched to assume that a history of a speech community with many of these "generalized outsiders", effects in the end that the language adapts to the outsiders' preferences. For instance, we expect that the history of intense language contact between coastal urban Norwegian in the late Middle Ages and North-Germanic dialects has had a *leveling,* or *pidginising*, or *koineising*, that is, *simplifying* effect on Norwegian.

As we measure complexity relative to the generalized outsider, we expect that in the end the measure of complexity of full languages (E-languages), correlates with the amount of outsiders in the speech community history. To speak more smoothly about the speech communities, I propose that there are two extremes of speech communities, Type 1 and Type 2, of which the latter is an extreme "generalized outsider community". We may further characterize these communities as follows:[4]

The population of an idealized Type 1 community is relatively small, and most people know each other. Most interactions take place among members of the community,

4. Note that the various aspects of the two types (e.g., size, kind of network structure, amount of contact with outsiders) are not independent but mutually reinforcing. This is a disadvantage from an analytical point of view. It is however an advantage with respect to its description from

and in interactions with outsiders a different language is used. Members of a Type 1 community share a large common background. Life cycles are relatively predictable, originality and innovation are not appreciated very much, and neither is the transmission of new information. Language is used to keep social relations between members of the community in balance. There is a body of literature (written or oral) in the community with a sacrosanct status. People are proud of their language, and they have stories that relate the origin and form of the language to their religion and their cultural origins. In everyday life, verbal play and language skills are appreciated. Different registers exist but there is no dialectal variation. Ties between community members and the possession of communal values are strong and no local centers of prestige develop, and therefore, also no dialects.

Type 2 communities only form a speech community because all members use the same language. They do not necessarily form a unit in space or time, or share social or cultural values. The number of speakers of a Type 2 language can be high. Most speakers know other languages as well, and the language in question is often not their first language, but only used as a lingua franca. The members of such communities do not share much background knowledge, and their precise way of speaking the language may differ. In interactions the language is mostly used for negotiating and exchanging practical information. The language is not associated with a cultural or religious standard, and speakers accommodate their way of speech freely in order to be clearly understood. The language is not a medium in which identities are expressed. There are no registers, but there may be different dialects, or different ways of speaking, possibly related to the original languages of the speakers.

In order to examine complexity differences most fruitfully it is easiest to compare related languages, and related speech communities, which have a common mother language, and only a different social history. If so, we may conjecture that when a language splits, and one variety becomes more like a Type 1, and the other like a Type 2 community, we expect that the latter becomes simpler in its inflectional morphology.

5. Case study: Bolivian, Argentinean and Ecuadorean Quechua

5.1 Social history of three Quechua varieties

I will now show how this turns out for one case of inflectional changes in related varieties: the Quechua languages of Peru, Argentina and Ecuador. Instead of expounding a wealth of tiny examples from diverse language families, I concentrate on

an interpretive point of view. Compare also the traits of a "personality type" in classical psychology, which are not mutually independent, though mutually reinforcing the saliency of the type.

one particular case. This provides the opportunity to search for the precise linguistic locations of complexity; to follow the exact paths of simplification among related varieties, and to find out how the actual implementation of the theory fares. In addition, by focusing on Quechua, we will find out that simplification can mean something quite different from the Indo-European based notion of "erosion", that commonly occurs at word boundaries. I will first describe some features of the histories of the Quechua varieties, and next compare their morphologies (for a more comprehensive treatment, cf. Kusters, 2003: 249–303 and references therein).

The origin of Quechua lies on the coast of central Peru before 500AD. Some varieties, called Quechua II, spread slowly southward along the coast, while Quechua I moved over the highlands. There was a second split before 1500, when traders moved northwards along the coast as far as Ecuador and used Quechua as a lingua franca (which became known as Quechua IIb or Ecuadorian Quechua). Meanwhile Quechua IIc was spoken near Ayacucho and Cuzco, which would grow in power as the capital of the Incas. Quechua spread to Bolivia and Argentina under the Incas, while it was learned in special schools, in ethnic interactions and in official Inca religious and administrative transactions. After the Spanish Conquest in the 1530s the status of Quechua changed. Initially the Spanish government promoted Quechua, but later in the 18th century Quechua was associated with backwardness and its existence came under threat.

In the first decades after the Conquest, Quechua was used as the lingua franca in southern Peru by Andean migrants for communicative purposes and as a church language. Later Spanish became more important. Today, in spite of centuries of repression of the Andean culture and extensive migrations and depopulations, Quechua is still widely spoken. Cuzco Quechua has become the prestige language – both spoken and written – for an indigenous upper class of Andean nobles and it is supported by a language academy. We expect some early simplification in Cuzco and Ayacucho varieties, because of the turbulent history of migration and population changes. In comparison with the other Quechua varieties, we expect less inflectional change, since in Peru it was not adopted as a second language or lingua franca by large groups of L2 learners.

After arrival in Argentina under the Incas Quechua further spread and was consolidated under the Spaniards. The heterogeneity of speakers who took Quechua there was much higher than in Peru. First and second language speakers of Quechua, Andean migrant workers from diverse Quechua varieties and other Andean languages, Spanish colonists and Spanish priests were all involved in its spread and development. Argentinean Quechua became a mix of different native, koineized and second language varieties of Quechua. It does not have a written literary tradition, and has far less status than Spanish. Because of the quite rapid L2 acquisition process, the wide diversity of speaker backgrounds involved in its early spread, and its continuous low prestige, we assess the Argentinean Quechua speech community as far closer to a Type 2 community than the one in Peru.

In pre-Inca times Quechua was adopted in Ecuador as a trade language, while the second Quechua wave came through the Inca expansion, by which Inca Quechua

spread over the first layer of Quechua IIb speakers, and gained new Andean L2 speakers from various backgrounds. After an initial decrease due to the decline in population during and after the Inca civil war Quechua began to extend its domain as the indigenous lingua franca that expressed Andean identity. Among the Quechua communities discussed here, the Ecuadorian one displays most traits of a Type 2 speech community because of its early history as a trade language.

5.2 Quechua inflection

The total number of categories in a word is high in Quechua, while the number expressed in each affix is low. The inflectional categories on which I focus are tense, subject agreement, object agreement and number. The general order in verbal morphology in all Quechua II varieties is: VERB – OBJ -TENSE – SUB – NUM. For example, in Cuzco Quechua:

(4) *Yanapa-wa-rqa-nki-cis.*
help-1-PAST-2-PL(EXCL)
'You (plur) have helped me.'

Quechua varieties deviate in different ways from this template. Some categories appear in fused forms, sometimes the order deviates, and sometimes the meaning of the affixes is dependent on the meaning of other affixes.

In the next tables, I show the past tense paradigms of Cuzco, Ayacucho (in italics when different from Cuzco), Argentinean and Ecuadorian Quechua. Each column refers to a different object, while each row refers to a different subject. The endings which these different subject-object combinations display are spelled out in the cells.

Table 2. Cuzco and Ayacucho Quechua past tense inflection

	1SG	2SG	1PL.INCL	1PL.EXCL	2PL	3/NO OBJ
1 SG	*	rqa-yki	*	*	rqa-yki-cis	rqa-ni
2 SG	wa-rqa-nki	*	*	wa-rqa-nki-ku	*	rqa-nki
3 SG	wa-rqa-n	rqa-sunki *su-rqa-nki*	wa-rqa-ncis	wa-rqa-n-ku	rqa-sunki-cis *su-rqa-nki-cik*	rqa
1PL.INCL	*	*	*	*	*	rqa-ncis
1PL.EXCL	*	rqa-yki-ku	*	*	rqa-yki-ku	rqa-y-ku *rqa-ni-ku*
2PL	wa-rqa-nki-cis	*	*	wa-rqa-nki-ku	*	rqa-nki-cis
3PL	wa-rqa-n-ku *wa-rqa-n*	rqa-sunki-ku *su-rqa-nki*	wa-rqa-ncis	wa-rqa-n-ku	rqa-sunki-cis *su-rqa-nki-cik*	rqa-ku

Cuzco and Ayacucho Quechua differ with respect to the possibilities of the plural 3rd person marker, the 1st person plural marker, and the position of the *su*-affix (i.e., its fusion with the *nki*-affix).

Table 3. Argentinean Quechua past tense inflection (differences with Peruvian Quechua are shaded).

	1SG	2SG	1PL.INCL	1PL.EXCL	2PL	3/NO OBJ
1SG	*	su-ra-ni/ ra-yki	*	*	ra-yki-cis	ra-ni
2SG	a-ra-nki	*	*	a-ra-yku	*	ra-nki
3SG	a-ra	su-ra	a-ra-ncis	a-ra-yku	su-ra-nki-cis	ra
1PL.INCL	*	*	*	*	*	ra-ncis
1PL.EXCL	*	su-ra-yku	*	*	ra-yki-cis	ra-yku
2PL	a-ra-nki-cis	*	*	a-ra-yku	*	ra-nki-cis
3PL	a-ra-nku	su-ra-nku	a-ra-ncis	a-ra-yku	su-ra-nki-cis	ra-nku

Apart from phonological differences, Argentinean differs with respect to: (1) the reanalysis of *su*, this affix has turned into a 2nd person object marker, (2) the reanalysis of *y-ku* and *n-ku* into fused morphemes, which prevents separate *ku* from appearing. As a result of these minimal changes, maximal paradigmatic effects occur (cf. Kusters 2003: 276ff.). Most notable are the less transparent ways to express 1st person plural exclusive objects, and the more transparent way to express a 2nd person object.

As discussed in Section 3, it is not completely straightforward how to assess the changes in such paradigms, but taking everything into account our impression is that this paradigm is a little simpler than the Peruvian varieties.

Table 4. Ecuadorian Quechua past tense inflection

	1SG	2SG	1PL.INCL	1PL.EXCL	2PL	3/NO OBJ
1SG	*	*	*	*	*	-rka-ni
2SG	-wa-rka-ngi	*	*	*	*	-rka-ngi
3SG	-wa-rka	*	*	*	*	-rka
1PL.INCL	*	*	*	*	*	-rka-nci
1PL.EXCL	*	*	*	*	*	-rka-nci
2PL	-wa-rka-ngi-ci	*	*	*	*	-rka-ngi-ci
3PL	-wa-rka-(kuna)	*	*	*	*	-rka-(kuna)

This paradigm is univocally simpler than the Peruvian ones, on all three morphological dimensions of Section 5.1. Affixes that do not express a clear unambiguous meaning (*yki, sunki*) are lost, and in addition, the plural marker has a more local scope: it can only pluralize its adjacent affix.

5.3 Conclusions and interpretations

We expected Cuzco Quechua to be more like a Type 1 language (that is, language characteristic for a Type 1 community) than Argentinean Quechua, and Ecuadorian

Quechua to be most like a Type 2 language. We also expected that all Quechua varieties would display Type 2 language characteristics because of the turbulent Andean history both under the Incas and after the Spanish Conquest.

The complex morphology of Cuzco and Ayacucho Quechua has, however, not essentially changed. How to explain this? This conservatism may be a reaction of the Quechua speakers to outside pressure exerted on their culture and language. Perhaps the new learners of Quechua from other Indian communities could not exert their influence on Quechua, because of the prestige of Quechua or because of the possibly relatively low numbers of learners. Furthermore, perhaps the stability on household level and intra-generational language transmission remained fully intact. However, perhaps the clear difference in the extent of simplification between e.g., Scandinavian and Quechua can be explained by the agglutinative structure of Quechua verbs, which may be intrinsically more stable than Old Norse structure (cf. Kusters 2003 for a far more extensive discussion).

Ecuadorian Quechua simplification can be explained by the following factors: (1) Unlike other varieties Ecuadorian Quechua was heavily influenced by traders who used Quechua as a lingua franca. Quechua was then acquired and used by adults who were not corrected by a strong Quechua norm. (2) The L2 learners of Quechua in Ecuador had probably more diverse and more deviant language structures than those in Argentina and certainly in Peru. In other cases of simplification as in Swahili, L1 similarity also determines the extent of simplification. Here we see the relevance of various senses of "relative complexity". A "generalized outsider" may have problems with Quechua that the actual Quechua learners in Peru did not have, because of their L1 background.

In Argentina, *su* is on its way to becoming a transparent object marker. This extension and regularization of object agreement may be an autonomous change, but because of its exceptionality in Quechua varieties, it counts as a simplifying change. As far as we may speak then of Argentinean simplification, it is interpreted in terms of the diversity of language backgrounds in Argentina, and the little contact Argentina had with the Quechua heartland.

However, Argentinean Quechua is less Transparent and Economical than Ecuadorian Quechua in its verbal inflection, although both countries seem to display similar social historical patterns. A plausible explanation is that all speakers of Ecuadorian Quechua were initially second language learners, while in Argentina there was a core of native speakers, who could function as a norm for "correct" Quechua. Argentinean Quechua served as a means of inter-ethnic communication and was not used exclusively for trading purposes or only by foreign language learners.

From studies in Arabic and Scandinavian we know that simplification may also be analysed as incidental effects of language-internal changes, like stress shift and concomitant vowel reduction and loss of affix distinctions. Such language-internal explanations are implausible for Quechua. The Quechua affixes vanished quite abruptly, and there were no intermediate stages in which there was a phonologically reduced form. Moreover, the affixes that disappeared in Quechua were not word-final, and they disappeared without

any other stress shift. Cuzco Quechua even contains a shift in metrical structure which is similar to that of Germanic, but still did not erode at the end of the word.

6. Conclusions

In spite of the controversies of the notion, this article – as well as the other articles in this volume – shows that "complexity" is a viable notion in linguistics. When properly defined, there are no a priori arguments why we could not profit from this concept. The issue whether complexity is best used as an absolute or a relative notion, is complex itself. I have argued that when we base our idea of complexity on the experiences of a particular type of language user, we receive more insights from subfields like theoretical linguistics, creole studies, contact linguistics, historical linguistics and typology.

When we apply the concept of complexity to an actual case of language change – in this article, change in Quechua verbal inflection – we may conclude that languages are adapted in their level of complexity to the preferences of language users. We also find that these preferences do not immediately translate in simplifications but are mediated by the characteristics of the structure of the language in question. The case of Quechua is interesting in this respect, because it shows what kinds of simplification are possible besides the well-known cases of Indo-European language change, which are all too often plainly characterized as "erosion". While "erosion" brings notions of natural (i.e., causal) change with it, the kind of simplification in Quechua invokes ideas of teleology and goal-directedness in matters of language change.

As usual in models that take social and historical facts into account, predictions do not come out as clear-cut as we might wish. Broad generalizations have all kinds of stubborn contours at closer look. Nevertheless, when we further examine the relevant social-historical factors and the morphological structures and confront these with empirical data, we may arrive at a new typology of simplification paths in the future.

Abbreviations

1	first person
2	second person
3	third person
EXCL	exclusive
INCL	inclusive
OBJ	object
PL	plural
PST	past
SG	singular

References

Aitchison, J. 1991. *Language Change: Progress or Decay?* Cambridge: Cambridge University Press.

Andersen, H. 1988. Center and periphery: Adoption, diffusion and spread. In *Historical Dialectology: Regional and Social*, J. Fisiak (ed.), 39–83. Berlin: Mouton de Gruyter.

Braunmüller, K. 1984. Morphologische Undurchsichtigkeit – ein Charakteristicum kleiner Sprachen. *Kopenhagener Beiträge zur germanistischen Linguistik* 22: 48–68.

Braunmüller, K. 1990. Komplexe flexionssysteme – (k)ein Problem für die Natürlichkeitstheorie? *Zeitschrift für die phonetische Sprachwissenschaft und Kommunikationsforschung* 43: 625–635.

Dahl, Ö. 2004. *The Growth and Maintenance of Linguistic Complexity.* Amsterdam: John Benjamins.

Efron, B. & Thisted, R. 1976. Estimating the number of unknown species: How many words did Shakespeare know? *Biometrika*, 63 (3): 435–437.

Gadelii, K.E. 1997. *Lesser Antillean French Creole and Universal Grammar* [Gothenburg Monographs in Linguistics 15]. Gothenburg: Gothenburg University.

Hyams, N. 1987. The Theory of Parameter and Syntactic Development. In *Parameter Setting*, T. Roeper & E. Williams (eds), 91–121. Dordrecht: Reidel.

Kerke, S.C. van de 1996. Affix order and interpretation in Bolivian Quechua. PhD Dissertation, University of Amsterdam.

Klein, W. & Perdue, C. 1997. The basic variety (or: Couldn't natural languages be much simpler?). *Second Language Research* 13: 301–347.

Kusters, W. 2003. Linguistic Complexity; the Influence of Social Change on Verbal Inflection. PhD Dissertation, University of Leiden.

Lakämper, R. & Wunderlich, D. 1998. Person marking in Quechua: A constraint-based minimalist analysis. *Lingua* 105: 113–148.

Maddieson, I. 1984. *Patterns of Sound.* Cambridge: Cambridge University Press.

McWhorter, J. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5: 125–166.

Miestamo, M. 2006. On the feasibility of complexity metrics. In *FinEst Linguistics, Proceedings of the Annual Finnish and Estonian Conference of Linguistics, Tallinn, May 6–7, 2004* [Publications of the Department of Estonian of Tallinn University 8], K. Kerge and M.-M. Sepper (eds), 11–26. Tallinn: TLÜ.

Mühlhäusler, P. 1974. *Pidginization and Simplification of Language.* Canberra: Linguistic Circle.

Nagel, T. 1986. *The View from Nowhere.* Oxford: Oxford University Press.

Sapir, E. 1921. *Language.* London: Harcourt, Brace & World, Inc.

Thurston, W.R. 1982. *A Comparative Study of Anem and Lusi.* Toronto: University of Toronto.

Thurston, W.R. 1987. *Processes of Change in the Languages of North-Western New Britain* [Pacific Linguistics B 99]. Canberra: Australian National University.

Thurston, W.R. 1992. Sociolinguistic typology and other factors effecting change in north-western New Britain, Papua New Guinea. In *Culture Change, Language Change; Case Studies from Melanesia* [Pacific Linguistics C 120], T. Dutton (ed.), 123–139. Canberra: Australian National University.

Trudgill, P. 1986. *Dialects in Contact.* Oxford: Blackwell.

Trudgill, P. 1992. Dialect typology and social structure. In *Language Contact, Theoretical and Empirical Studies*, E.H. Jahr (ed.), 195–211. Berlin: Mouton de Gruyter.

Trudgill, P. 1996. Dialect typology: Isolation, social network and phonological structure. In *Festschrift for William Labov*, G. Guy (ed.), 3–22. Amsterdam: John Benjamins.

Trudgill, P. 1997. Typology and sociolinguistics: Linguistic structure, social structure and explanatory comparative dialectology. *Folia Linguistica* 31: 349–360.

Trudgill, P. 2001. Contact and simplification: Historical baggage and directionality in linguistic change. *Linguistic Typology* 5: 371–374.

Winford, D. 1998. Creoles in the Context of Contact Linguistics. Paper presented at the symposium Pidgin and creole linguistics in the 21st century, New York, January 1998.