# The transparency of creoles*

Sterre Leufkens
Universiteit van Amsterdam

In this article I propose that creoles are relatively transparent compared to their source languages. This means that they display more one-to-one relations between meaning and form. Transparency should be distinguished from the concepts of simplicity, ease of acquisition, and regularity. Definitions of these notions are given and it is shown how they have been mixed up in earlier literature.

The transparency of creoles is explained as a result of language contact. When people speaking radically different languages communicate, they tend to use maximally intelligible forms, i.e. transparent forms. The repeated selection of transparent over opaque forms will lead to the formation of a relatively transparent language. Hence, creoles are predicted to be either as transparent as or more transparent than their source languages.

An empirical study is performed to test this prediction. The transparency of four contact languages and their sub- and superstrates is measured by checking them on a list of non-transparent features. It turns out that they all exhibit opaque structures, but that there is a striking absence of so called form-based forms: linguistic elements and rules that are not motivated pragmatically or semantically. This indicates that such 'empty' forms are lost during intense language contact.

**Keywords:** creoles, (semantic) transparency, simplicity, regularity, language contact, typology, language acquisition

* I am greatly indebted to Kees Hengeveld, Norval Smith, John McWhorter, Peter Bakker, and an anonymous reviewer for their accurate feedback and thoughts on earlier versions of this paper. The participants of the *Creole Grammars and Linguistic Theories* conference in Paris, June 13–14 2011 helped me enormously as well with their questions and comments. I also want to express my gratitude to those who helped me in analyzing the languages of their expertise: Enoch Aboh, Hugo Cardoso, James Gair, Mohamed Jaffar, Evelien Keizer, Lachlan Mackenzie, Sebastian Nordhoff, Caroline Roset, Ian Smith, and Joceli Stassi Sé.

*Journal of Pidgin and Creole Languages* 28:2 (2013), **323–362**. DOI 10.1075/jpcl.28.2.03leu
ISSN 0920–9034 / E-ISSN 1569–9870 

## 1. Introduction

It has been claimed that creoles are simpler, in certain respects, than non-creole languages (notably by McWhorter 2001 and more recently by Parkvall 2008). In the current article, I will argue that creoles are not simpler but more transparent than non-creoles. Transparency is defined as a one-to-one relation between meaning and form and is not the same as simplicity. Transparency is a non-discrete notion; creoles have a relatively high degree of transparency compared to non-creole languages. It is not impossible for a non-creole to be highly transparent. Therefore, it is not possible to categorize a language as a creole or a non-creole on the basis of its degree of transparency alone.

The term transparency has been used before in theoretical explanations of language change, for example by Langacker (1977), and in the realm of language acquisition by Slobin (1977). The term relates to but is not identical to semantic transparency as defined by Seuren & Wekker (1986) — the first linguists to introduce the term in creole studies (cf. Arends et al. 1995: 11). Ever since the conception of the term, transparency and simplicity have often been mixed up, which has led to confusion about which characteristic(s) exactly define(s) creoles. In my opinion, the term 'simplicity' is often used so vaguely that it cannot be a tool in understanding the characteristics of creoles. I hope to contribute to the ongoing simplicity debate by clarifying what exactly this and other terms mean.

The paper is structured as follows. Section 2 gives an account of the difference between transparency and simplicity. These and other relevant terms are carefully defined, in order to separate the concepts involved. Section 3 is devoted to contact languages. It is argued that languages become more transparent under the pressure of language contact. In Section 4, I will explain in greater detail what transparency is and how it can be measured. Following that, an empirical study will be presented in Section 5. In this study, the transparency of four contact languages and their sub- and superstrate languages will be measured and compared. The conclusions of the study will be given in Section 6, which also provides suggestions for further research.

## 2. Transparency vs. simplicity, ease of acquisition, and regularity

### 2.1 (Semantic) transparency and simplicity

In this section I will show how the notion of 'transparency' has been confused with the notion 'linguistic simplicity', a term that itself has various interpretations. Subsequently, I will try to establish a clear terminology by defining four concepts

(transparency, simplicity, ease of acquisition, and regularity) separately. This is essential for a proper understanding of the phenomena at hand and consequently for understanding contact languages. Finally, I will show how the different notions relate to each other.

The term transparency has been interpreted in various different ways over the years. In my opinion, a transparent relation is a one-to-one relation between meaning and form (Hengeveld 2011), where a form can be a morpheme, word, or a larger linguistic unit. We find numerous structures in languages of the world that violate this principle: so called non-transparent or opaque (I will use these terms interchangeably) structures. A fully transparent language is a language without any such violations. Such a language does probably not exist — every language violates transparency somewhere in its grammar. However, languages can exhibit different degrees of (non-)transparency.

A different interpretation of the term transparency can be found in McWhorter (2001). In this article, McWhorter criticizes the so-called semantic transparency hypothesis, proposed by Seuren & Wekker (1986). This hypothesis comprises firstly the idea that creole languages are more transparent than other languages, and secondly that this supposed transparency is due to the circumstances in which creoles emerged. Semantic transparency[1] involves the maximization of several principles, one of which is 'uniformity of treatment of semantic categories' (Seuren & Wekker 1986: 64), which is the one-to-one meaning-form principle referred to above.

McWhorter, however, follows Kihm (2000) in interpreting semantic transparency as semantic atomicity: the degree to which semantic atoms are expressed as separate lexical items, rather than in 'unitary equivalents' (McWhorter 2001: 156). The English lexeme *to fetch* is for example a unitary equivalent of the semantic atoms GO, TAKE, and COME. By unifying semantic units in one lexeme, the English word is non-transparent. Showing that Vietnamese, a non-creole language, is very 'atomistic', McWhorter argues that transparency cannot be *the* defining characteristic of creoles. According to him, there might be a tendency that creoles are relatively transparent (Kihm 2000: 186 notes this as well), but they cannot be defined on the basis of transparency alone.[2]

1. Semantic transparency is not exactly the same as what I call transparency. It involves, apart from a one-to-one meaning-form relation, minimal processing and a minimal amount of language-particular rules (Seuren & Wekker, 1986: 64). Transparency in my definition only comprises the one-to-one meaning-to-form relation.

2. In fact, I agree with McWhorter that transparency is not the characteristic on the basis of which one can distinguish creoles from non-creoles. Rather, I think that creoles group more to the transparent end of the spectrum, assuming a continuum between transparency and opacity.

Without pursuing McWhorter's ideas any further here, we can safely say that his interpretation of semantic transparency is different from the one that Seuren & Wekker propose. They do not mention semantic atoms; in fact, they argue that the notion of 'semantic element' is a highly problematic one (Seuren & Wekker 1986: 63), for instance because determining what exactly a semantic element is (an atom, in McWhorter's terms) is very much a theory-dependent decision.[3]

Furthermore, McWhorter's (2001) main point is that creole languages are *simpler* than non-creole languages, according to a simplicity measure (adopted in Parkvall 2008) that involves counting the number of overt distinctions and forms in different areas of grammar (i.e. the number of marked phonemes, of syntactic rules, of expressed semantic distinctions and the amount of complex inflectional morphology). According to McWhorter, linguistic complexity furthermore increases with the occurrence of certain morphological phenomena: suppletion, allomorphy (e.g. declensional classes), and agreement. In McWhorter (1998), irregularity in derivation and compounding are listed as creole characteristics, but these features do not reoccur in the 2001 article as complexifying factors.

In sum, McWhorter argues against the idea that semantic transparency is a typical feature (let alone a defining characteristic) of creoles. Nonetheless, in many articles the ideas of McWhorter are equated with the semantic transparency hypothesis. Braun & Plag (2003), for instance, interpret this hypothesis as follows:

> In Thomason's words (2001: 168), '[m]orphology also tends to be extremely regular when it does exist in pidgins and creoles, without the widespread irregularities that are so very common (to the distress of students of foreign languages) in other languages' morphological systems'. In what follows, we will call this 'the semantic transparency hypothesis' (Braun & Plag 2003: 81).

Braun & Plag (2003: 81) then write, 'This hypothesis is explicitly argued for by Seuren & Wekker (1986) and, in considerable detail, more recently by McWhorter (1998, 2001)'. The authors do not take McWhorter's (2001) explicit rejection of the transparency hypothesis into consideration and equate McWhorter's ideas on *simplicity* with it. They proceed by presenting examples of irregular morphology in creoles, thus rejecting what they believe is the transparency hypothesis. Of course,

However, where McWhorter believes that simplicity does constitute a defining creole characteristic, I believe that there is no linguistic feature that distinguishes creoles from non-creoles (cf. Section 3).

**3.** As a reviewer points out, different authors need not agree on how a notion should be defined. Moreover, Seuren & Wekker do not have an *a priori* correct definition of transparency, just because they were the first to introduce it in creole studies. What I want to show here is that the notion of transparency has gotten *different* interpretations over the years, not that one of the interpretations is the one and only correct one.

what is actually rejected here is McWhorter's (1998) hypothesis that morphological regularity is a characteristic of creoles. Seuren & Wekker (1986) never made this claim. Moreover, no such claim follows from the hypothesis that creoles adhere more to a one-to-one relation between meaning and form.

The mix-up between supposed simplicity of creoles and the semantic transparency hypothesis is also apparent in Aboh & Smith (2009). They write:

> On the morphological level, simplicity is taken to mean that creole languages display a rather poor morphology, whether inflectional or derivational, and mainly resort to compounding […]. The rationale here is that such a lack of morphology would favor a semantic transparency not found in other natural languages (Aboh & Smith 2009: 3).

Semantic transparency is seen as the same as (or at least a consequence of) simplicity and a poor inflectional or derivational morphology. The same confusion is manifested in, for instance, Lefebvre (2001) and Kihm (2000: 176), who remarks that '[…] studies conducted in the "simplicity" or "semantic transparency" paradigm have had a tendency to focus on the verb phrase'. Kihm's use of the singular 'paradigm' indicates that he regards the two concepts as identical.

**Table 1.** Two lists of 'complex' features

| **McWhorter (2001: 163) — Features never found in creoles** | **Dahl (2004: 115) — Maturation phenomena** |
|---|---|
| Ergativity | Inflectional morphology |
| Grammaticalized evidential marking | Derivational morphology |
| Inalienable possessive marking | Incorporating constructions |
| Switch-reference marking | → Grammatical gender |
| Obviative marking | → Inflectional classes |
| → 'Dummy' verbs | → Idiosyncratic case marking |
| → Syntactic asymmetries between matrix and subordinate clauses | → Agreement |
| Grammaticalized subjunctive marking | → Word order rules over and above internal ordering of sister constituents |
| → Verb-second | → Specific marking of subordinate clauses |
| Clitic movement | → Morpheme and word level features in phonology |
| Any pragmatically neutral word order but SVO | |
| → Noun class or grammatical gender marking | |
| Lexically contrastive or morphosyntactic tone | |

There are two reasons for this terminological confusion. First of all, the theories of Seuren & Wekker and McWhorter both assume that creoles are a distinct type of language. The only difference is the nature of the feature(s) discriminating them from non-creoles. Secondly, assuming simplicity or transparency to be the main discriminating feature results in roughly similar lists of items that are unexpected in creoles. For example McWhorter's (2001: 163) list of features never found in creoles contains many features that also violate a one-to-one meaning-to-form relation, and so does Dahl's (2004: 115) list of 'maturation phenomena' (complexifying properties that languages supposedly acquire over time). These lists are given in Table 1. The features that violate transparency (according to the list proposed in Section 4 below) are marked with →. Given the high amount of arrows, it is perhaps not surprising that accounts of simplicity are mixed up with accounts of transparency.

### 2.2 Disentangling transparency, simplicity, ease of acquisition, and regularity

To be able to study the characteristics of creoles, it is necessary to determine what they are, and how they differ from the characteristics of non-creoles. The aim here will not be to provide extensive and precise definitions of all the relevant concepts, but to characterize them as far as necessary to render them distinguishable. The first, **transparency**, was defined above as a consistent one-to-one relation between meaning and form. An overview of transparent linguistic features will be given in Section 4. Other influential definitions of transparency can be found in Langacker (1977), Seuren & Wekker (1986), McWhorter (1998, 2001) and Kusters (2003). These accounts will be discussed at the appropriate places.

Linguistic **simplicity** should in my opinion be defined in terms of a small amount and a relative superficiality (i.e. few layers of embedding) of linguistic material. In other words: the less linguistic material is used (for any given message) in a language and the more superficial its structure, the more simple the language is. So a sentence with fewer morphemes is relatively simple (surface simplicity), and a sentence with fewer embedded phrases or clauses is also relatively simple (structure simplicity). This kind of simplicity, which does not refer to acquisition but is in fact an information-theoretic notion (Dahl 2004: 39), has been called absolute simplicity (Miestamo 2006).

McWhorter's (2001) simplicity metric, referred to above, is an example of a definition of absolute simplicity. Another one is from Langacker (1977), who sees simplicity as a combination of signal simplicity ('fewer and shorter units of expression', p. 112), perceptual optimality (saliency), constructional simplicity (syntactic depth of linguistic material) and transparency. Similarly, Dahl (2004) splits linguistic simplicity in separate notions such as structural simplicity (a low

amount of material at some level of organization) and system simplicity (simplicity of the mappings from meaning to form, p. 43 — transparency is therefore one factor in Dahl's simplicity measure). The overall degree of simplicity of a language is, in both Langacker's and Dahl's accounts, the total of these separate measures. Features that are considered simple in these approaches are for instance:

- A relatively small amount of morphological and phonological rules
- A relatively small phoneme inventory
- Shallow structure, i.e. a relatively small degree of syntactic depth, e.g. no subordination but only coordination
- Regularity (exceptions to paradigms are extra linguistic material and add complexity)

As Aboh & Smith (2009) point out, measures of absolute complexity only deal with overt marking, being unclear about the interaction between overtness and covertness and about how covert markers might add to complexity.

Some authors define simplicity in a different way, referring to **ease of acquisition**. For instance Kusters (2003) states that a language is simpler if it is easier to acquire for an L2 learner. Simplicity in this sense is called 'relative simplicity' (Miestamo 2006). For reasons of clarity, I prefer not to use the term simplicity in this sense. I think it is important to keep apart the simplicity of a language system as such and the simplicity of *the acquisition* of a language system, as these are two different things. They should ideally be studied separately, as they might not always go hand in hand: as will be argued below, a simple system is not necessarily easier to acquire. Referring to both 'simplicities' with one word mixes up properties of the system itself and properties of humans learning the system. I will therefore use 'simplicity' only for absolute simplicity, and reserve the term 'ease of acquisition' for relative simplicity.

Properties that make a language or feature easier to acquire are for instance:

- In the case of an L2 learner: a smaller difference between L1 and L2
- High perceptual salience (of a feature)
- High frequency (of a feature)
- Regularity (exceptions to rules have to be learned in addition to rules themselves)
- Transparency (see below)

A fourth notion that has been mixed up with transparency is **regularity**. Regularity can refer to the predictability of paradigms, that is, the way in which a paradigm is consistently followed by all members of the set that the paradigm applies to. I will call this **paradigmatic regularity**. The past tense inflection of verbs in Germanic languages is an example of paradigmatic irregularity.

In another sense, regularity refers to the predictability of meaning from its form. As the relation between form and meaning is in most cases arbitrary (with the exception of onomatopoeia), this is not possible for most forms (one cannot predict what e.g. 'green' means by knowing the visual, auditory or orthographic form of the word). But the matter becomes more relevant when we look at combinations of forms, for instance in syntax or in compounds and derivations. The latter is actually what McWhorter (1998) is referring to when he speaks of **semantic regularity** or derivational regularity: a derivation or compound is semantically regular if its meaning can be predicted from the meanings of the combined forms (cf. Kouwenberg & LaCharité 2011, who relate regularity to compositionality). For example, the compound noun 'lipstick' is considered to be semantically regular, because its meaning is composed of the respective meanings of 'lip' and 'stick'. A derivation like 'computer' is semantically irregular, as it is not a straightforward combination of the meanings of its separate parts: a computer is not only 'someone or something that performs computations' — a computer does more.

Aboh & Smith (2009) argue that in fact all derivations and compounds are semantically irregular. There are two reasons for this. First of all, in the meaning of a compound or derivation only *a part of* the meanings of the separate elements is represented. One cannot predict which part this is. When someone hears the word *lipstick* for the first time, (s)he might think that it refers to 'a stick made of lips'. That interpretation includes the semantics of *lip* and *stick*, but not the right aspects (i.e. the body part and the cylindrical shape).[4] Secondly, the semantic relation between combined elements can never be predicted. Again, when someone hears a word like *lipstick* for the first time, (s)he cannot predict what it means on the basis of the meanings of *lip* and *stick*, because (s)he does not know how the two are related (is it a stick made of lips, a lip-shaped stick, something that sticks to lips?). A compound (or any other combination of elements) is only fully semantically regular when the semantic relation between the elements is in some way retrievable — for instance if in a particular language only part-whole relations could lead to compounds. There is to my knowledge no language where this is the case, so that all languages are at least to some extent non-transparent in this respect. Of course, there might very well be differences in degree here, but at this point I have no way to establish something like a 'scale of semantic irregularity'. I will therefore not include semantic irregularity in the list of opaque features given in Section 4.

4. Of course, this is due to the way that compounds and derivations come into being. First, there is some concept to be named. When there is a (strong) resemblance between this concept and some other concept, it is convenient to adopt the word for the other concept. That a word is used in a derivation or compound does not mean that the *complete* semantics of the word is adopted.

There is one more thing to be noted here. A compound is seen as an analysable unit here, that is, a composite of two elements. Of course, this is not necessarily the case for a language user — maybe a speaker or hearer just accesses a compound or derivation as one unit from her lexicon, not realizing it is composed of two elements, and without actively combining semantics. In this scenario, there is nothing interpretable, predictable, and hence nothing (ir)regular about it. This is very much a matter of productivity and lexicalization: is composition a productive process (i.e. are users actively compounding/deriving) or do they retrieve units ready-made? I will assume here that in every language there are at least some composite units that are analyzed, i.e. derived in use, so that there is irregular derivation in all languages.

### 2.3 (How) are transparency, simplicity, ease of acquisition and regularity related?

I have argued that in previous discussions on transparency and simplicity, we have been dealing with four different concepts: transparency (one-to-one meaning-form relation), simplicity (low amount and low degree of embedding of linguistic material), ease of acquisition, and (paradigmatic and semantic) regularity. Now that they are distinguished, it is possible to look at their interrelations.

Let us consider the L1 acquisition of Turkish, elaborately described in Aksu-Koç & Slobin (1985). Turkish inflectional morphology is strongly agglutinating and very regular. The language avoids homophonous forms, allowing a strong one-to-one relation between morphemes and semantic functions. Examples (1)–(3) (all taken from Aksu-Koç & Slobin 1985: 841) illustrate the Turkish morphological system.

(1) Getir-me-di-n.
bring-NEG-PST-2SG
'You didn't bring (it).'

(2) Ağla-dı-lar.
cry-PST-PL
'They cried.'

(3) Düz kon-ul-ur-sa.
straight put-PASS-AOR-COND
'If one puts it straight.'

Clearly, the Turkish morphological system is very transparent.[5] However, Turkish inflectional morphology is complex (in the absolute sense), as many functions are obligatorily expressed: there are many overt distinctions and many morphemes. The case of Turkish shows that transparency and simplicity are separate properties of (parts of) languages, which do not necessarily go hand in hand. Something that is simple does not need to be transparent, and vice versa.

Turkish inflectional morphology[6] turns out to be relatively easy to acquire, as Turkish children have fully acquired the nominal inflection system by two years of age, and verbal inflection before the age of three (Aksu-Koç & Slobin, 1985: 845). This shows that complexity does not necessarily lead to difficulty in acquisition. In other words: complex inflectional morphology is not difficult to acquire, at least for L1 learners, as long as it is transparent.[7]

The fact that the highly transparent Turkish inflectional morphology is easy to acquire suggests that transparent relations are easier than non-transparent ones (at least for children). This is confirmed by the error pattern of Turkish children: the few errors that remain after the age of three concern non-transparent embedding structures. Five-year-old Turkish children struggle, for instance, with specialized forms (e.g. participles and nominalizations) of embedded clauses that are not semantically driven (Aksu-Koç & Slobin 1985: 858), illustrated in (4).

Aksu-Koç & Slobin (1985: 860)

(4) Bizim ev-in ön-ün-e
1PL.GEN house-GEN front-GEN-DAT
gel-en kedi-ye benz-iyor-Ø.
come-SUBJ.REL cat-DAT resemble-PROG-3SG
'It looks like the cat that comes to the front of our house.'

In this sentence, the relative clause has a non-finite verb (*gelen*), while it would be a finite verb in a main clause. This difference is not semantically but syntactically driven. The verb form is therefore a form without a meaning correlate, which is opaque. The fact that Turkish children acquire them later than semantically driven

5. But note that Turkish vowel harmony (illustrated by the past tense morpheme *-dV-* in examples (1) and (2)) is a non-transparent feature — this involves formal changes without semantic motivation. This nicely illustrates that a language can be very transparent in one part of its grammar, but opaque in another.

6. I will disregard (Turkish) derivational morphology here as I am not familiar with research on its acquisition.

7. Note that L2 acquisition is a different story: Blom et al. (2008) find that complex inflectional morphology is difficult to acquire for adult L2 learners compared to L1 and L2 acquiring children.

forms confirms that transparent features are relatively easy to acquire, while non-transparent features are more difficult. In the words of Aksu-Koç & Slobin (1985: 855): 'Clarity of semantic mapping probably facilitates acquisition'.

Examples of the relatively late acquisition of non-transparent phenomena[8] are found in other languages as well. The highly opaque grammatical gender system of Dutch is, for instance, not mastered before the age of 7 (Blom et al. 2008). Dutch diminutives, characterized by allomorphy, are mastered by only 87% of 7-year old native speakers (Snow et al. 1980). Children learning Egyptian Arabic master irregular (suppletive and hence non-transparent) plural inflections for nouns and adjectives at age 6 or later (Omar 1973: 189). Different studies indicate therefore that at least L1 acquisition, transparent relations are learned more easily and earlier than opaque ones. Note that I do not deny that simple structure is in general easier to learn than complex structure. I simply say that transparency is another factor that should be taken into account in explaining the learnability or easiness of a structure.

Finally, we need to address the relationship between regularity and the other concepts discussed. All paradigmatic irregularities are opaque, as they involve additional forms to express the same meaning. Semantic irregularity is, as explained above, necessarily non-transparent as well. As such, it follows from the discussion above that irregularity is more difficult to acquire than regularity. This is confirmed, for instance, by the Egyptian Arabic data referred to above — irregular plural inflection of nouns is acquired relatively late.

Paradigmatic irregularity also enhances complexity, as exceptions in paradigms necessarily involve a greater number of overt forms. For instance, the irregular present tense inflection of the verb *to be* in English involves a number of extra forms, whereas the regular present tense verbal inflection in English involves only the suffix *-s*. There is no specific relation between semantic irregularity and complexity — the non-predictability of combined semantics does not lead to more overt forms.

## 3. Transparency in creoles

Now that I have delineated the relevant concepts, let me return to the issue of creoles. I propose that language contact leads to more transparency in a language. As creoles are the result of intense language contact, I hypothesize that creoles are either more transparent than or as transparent as non-creoles, but never more

**8.** It will become clear in Section 4 how the examples given here violate a one-to-one meaning-to-form relation.

opaque. Before addressing the reasoning behind this, let me go into my conception of the term 'creole'.

The distinction made between creoles and non-creoles is often said to be sociohistoric rather than linguistic. According to, for example, Arends et al. (1995: 6), creoles are a separate class from non-creoles because of their history, not because of linguistic characteristics. Others, for instance McWhorter (1998) and most recently Bakker et al. (2011), would say that creoles *can* be discriminated from non-creoles on the basis of their grammars. My position on this is that (1) language contact leads to more transparency, (2) in the case of creoles there is more intense contact, hence more 'transparenticization'. A discrete distinction between creoles and non-creoles can in my opinion not be drawn on the basis of transparency or any other linguistic feature: the difference is quantitative, not qualitative. There might very well be highly transparent non-creoles (for instance Riau Indonesian, addressed in the work of David Gil), as well as relatively non-transparent creoles.

This means that the term 'creole' will be used in this paper as a sociohistoric term: creoles are languages that arose from language contact during the colonial period. They might, but do not necessarily, originate from pidgins. They are a subset of the group of contact languages, i.e. languages that are formed during some form of language contact and are at least to some extent based on the languages in contact. As such, I include under the heading of contact languages also so-called mixed or intertwined languages (Arends et al. 1995: 41ff), which are languages that supposedly combine features from source languages without reduction or adaptations.

In the account of language contact and the emergence of creoles I give below, I will simplify matters considerably. To give a really thorough description of what happened during creolization, one would have to specify precisely who spoke what language when to whom. In other words, one would have to have information on population structure (where did people come from and when, how were they socially related, how did this change over time) and language use (what was the vernacular, which languages and varieties were present, which variety had high prestige, what motivations there were to learn other people's language). For most creoles, we simply do not have such information, which is one reason for this simplified account. Another reason is that I want to target transparency, abstracting away from other matters that, admittedly, must also have had a large influence on the formation of creoles.

Every speaker has several linguistic means at her disposal to transmit a message. The choice for a particular linguistic form (e.g. a lexeme or construction) depends on the balance between competing motivations. There is a motivation to be sparing, as a speaker will presumably want to spend as little effort as necessary to get her message across. This motivation, 'economy', triggers the use of

reduced forms. On the other hand, there is a motivation, 'intelligibility', to be understood, which favours the use of forms that are easy to perceive.[9] Different situations may demand the precedence of one motivation over others. For instance, when a speaker is in a hurry, it is important to speak quickly and economy will determine the linguistic forms used. But when one is speaking to someone with hearing loss, more salient and intelligible forms will take precedence over economic forms.

In situations where speakers of different languages want to communicate, but have no common language available, the need for intelligibility will be extremely high.[10] Speakers will use forms that are easily intelligible and refrain from using reduced forms. On the assumption that transparent forms (that is, forms that correlate in a one-to-one fashion to meaning) are highly intelligible, we can expect that in a language contact situation, transparent forms will be popular. Opacity is not helpful and therefore infrequent in such circumstances. The repeated selection of transparent forms over opaque ones leads to linguistic change: the language will become more transparent overall.

These ideas are not new — they fit nicely with what for instance Aboh & Ansaldo (2009: 47) envision when they say:

> […] it appears that morphemes that are semantically active are highly competitive and make it to the emerging language, regardless of whether they are of a derivational or inflectional type. On the contrary, morphemes that are semantically vacuous, or light, are less competitive and fade out of the Feature Pool.

The crucial point is: of all language features the source languages contribute, the more transparent ones (i.e. the semantically motivated units) will be selected into the contact language. Inheritance is constrained by transparency.

I expect this effect to be strongest in situations where the languages in contact are typologically most distant. When a speaker of German meets a speaker of Dutch, the resulting contact variety will not be so different from the native languages of the speakers — the degree of transparency will also not differ dramatically. But when a speaker of Hausa meets a speaker of Portuguese, the distance between the L1s of the speakers on the one hand and the new contact language

**9.** Intelligibility also favours redundancy, a non-transparent phenomenon. I will come back to this in Section 6.

**10.** Note that I am only discussing those cases of language contact where speakers want to communicate and be intelligible. Of course, there are also contact languages that have a function in forming the identity of a group, e.g. Verlan. Such languages are meant to be comprehensible for an in-group, but incomprehensible for outsiders. As a reviewer rightly claims, such languages will likely be less transparent than their source language.

on the other would be extremely large.[11] If every separate change leads to more transparency, a higher amount of changes will lead to a much higher degree of transparency.

Secondly, the amount of change (and the concurrent loss of opacity) will be higher when the need to communicate in the contact variety is higher. If two speakers of different languages can avoid speaking to each other or have the time to learn another language, they have no need to invent a new variety or use intelligible forms. But if they have to communicate, for instance because they want to trade, chances are higher that they will create a new variety, thereby selecting transparent forms.

When these factors (high typological distance between parent languages and high reliance on new contact variety) collide, the loss of opacity will be strongest. This leads to the hypothesis that creoles (and other languages resulting from contact) are more transparent than their sub- and superstrate languages. Isolated languages (languages that are spoken by small communities that are not often in contact with communities speaking other languages) will show more opacity compared to contact languages.[12]

This idea is not new. Trudgill (2004), for example, has made a similar claim that open communities have smaller phoneme inventories than closed communities — language contact would lead to a loss of phonemes (i.e. a loss of complexity, not opacity). He compares several Polynesian languages to prove his claim. Bakker (2004) replies with a quantitative study of other languages, in which he finds that open communities do *not* have significantly smaller phoneme inventories. However, Bakker compares unrelated languages, while Trudgill compares languages of the same family. To say that Iai has a higher number of phonemes then English does not tell us anything about the effect of language contact on a phoneme inventory, as the histories of both inventories are entirely different. The Polynesian communities are comparable in this respect, so that it makes more sense to compare them and their languages. Following Trudgill's approach, I will compare the *relative* transparency of *related* languages in the current study.

The proposal that languages in contact are relatively more transparent was addressed as well in a recent study by Lupyan & Dale (2010). 2,000 languages were

**11.** Cf. Aboh (2009), who states that linguistic differences between creoles and non-creoles are due to the large typological distance between source languages, rather than to a qualitatively different kind of language contact.

**12.** Note that the account of language contact and creolization given here is only about transparency and makes no predictions about simplicity, ease of acquisition or regularity of contact languages.

studied on 28 typological features.[13] Many of these features are found to correlate with the type of communities where those languages are spoken. Languages spoken in 'exoteric' communities, i.e. communities with a great amount of neighbouring communities (and hence a lot of language contact), are found to be more analytical than related, less exoteric languages. Furthermore, such languages tend to have less agreement, and they possess less idiosyncratic morphology. Lupyan & Dale (2010: 6) state:

> We found that the relationship between exotericity and increased form-meaning compositionality holds not only for specific linguistic features like tense and evidentiality, but is also supported by the observation that languages in the exoteric niche are more likely to be classified by typologists as being isolating rather than concatenative or fusional.

A similar correlation between exoteric communities and certain typological features is found by Kusters (2003). Kusters speaks of 'a stricter obedience to the Transparency Principle' (2003: 357) in communities with a relatively large proportion of bilinguals.

The account of language contact outlined here has led to the hypothesis that language contact leads to an increased degree of transparency. Creoles and other languages resulting from contact must then be more transparent than or as transparent as the languages from which they originate, but never more opaque. The sociolinguistic situation will favour the selection of transparent features over opaque features. This hypothesis will be tested below.

## 4. Transparency in language: transparent and opaque features

### 4.1 A list of non-transparent features

In this section, I will go further into the definition of transparency introduced above. I will demonstrate which linguistic phenomena are transparent, and which phenomena violate a one-to-one meaning-form relation. Examples are provided

**13.** The sample involves the languages described in the World Atlas of Language Structures (Dryer & Haspelmath 2011), available online at http://wals.info/. The WALS contains information on 2,000 languages, but it is not the case that all languages are studied for each feature — the number of languages studied per feature is considerably lower than 2,000. Lupyan & Dale seem to take their definition of what constitutes a language from the WALS. The boundary between varieties and languages might in some cases be controversial. For instance, similar varieties in Africa are often seen as separate languages, while Schweitzerdeutsch is analyzed as a variety of the German language, even though it is unintelligible for speakers of other German varieties. I want to thank Norval Smith for pointing this out to me.

from four languages that emerged during situations of intense language contact: Nubi, Sri Lanka Malay (henceforth SLM), Pichi, Diu Indo-Portuguese (henceforth DIP), and their sub- and superstrate languages. For more information on these languages and a motivation for this sample, see Section 5.1.

This list of non-transparent linguistic phenomena[14] will be employed in Section 5 to measure the degree of transparency of languages. It is an adapted version of the list composed by Hengeveld (2011), which was created by means of checking possible relations between meaning and form units as defined in Functional Discourse Grammar (henceforth FDG; Hengeveld & Mackenzie 2008).[15] Note that the list is not exhaustive: it captures many, but not all non-transparent phenomena. The opaque features are divided over three categories: redundancy, domain disintegration and form-based form.

The current paper is restricted to grammatical phenomena: the (transparency of the) lexicon is not studied here. Many-to-one relations between words and their meanings (e.g. homonymy and polysemy) are therefore not included on the list.

The degree of transparency of a language can be measured by counting the non-transparent features that it exhibits. Ideally, the features should be weighted and each feature should be assigned a value on a scale — this would avoid the crudity of saying, for example, that a language either does or does not exhibit fusion. However, at the moment, such modifications are still too complex to operationalize — refining the measure remains a task for the future.

### 4.2 Redundancy

The first type of opacity is redundancy; this involves constructions or sentences in which one pragmatic or semantic unit has more than one formal equivalent. For instance, a one-to-two meaning-to-form relation results when a semantic negation is expressed two or more times morphosyntactically. This is called **negative concord**. An example is given in (5).

14. According to the definition used here, a construction is transparent when it does not contain any many-to-one or zero-to-one structures. It would be rather impractical to give a list of transparent relations, as there is usually no name for the transparent relation but only for the violation. Therefore, the list contains non-transparent rather than transparent features.

15. The reason for using FDG is that it provides a very concrete modelling of the interfaces between different components of language: pragmatics, semantics, morphosyntax, and phonology. Other frameworks provide good models of specific components, but do not provide the means of studying interface relations specifically. As the readers of this paper will probably not be familiar with FDG, I have tried to make the list as theory-neutral as possible. Note, however, that that is never fully possible — readers working within different frameworks might find specific theoretical choices inappropriate or irrelevant.

Diu Indo-Portuguese (Cardoso 2009: 211)

(5) Nīge nã apĩŋ-o pex.
nobody NEG catch-PST fish
'Nobody caught fish.'

The two negating words ('nobody' and 'not') relate to one semantic negation. Note that concord is not the same as agreement in this paper: agreements concerns a morphosyntactic copying rule resulting in an empty copied element, while concord is the double expression of a semantic unit: both formal units have a semantic value (see below).

Other semantic information that is sometimes expressed multiple times is number. Redundancy occurs when a numeral and nominal marking both express plurality. Such **plural concord** occurs for instance in SLM, as illustrated in example (6).

Sri Lanka Malay (Nordhoff 2009: 243)

(6) Kandi=ka hathu thigapulu riibu=kee mlaayu pada arà-duuduk.
Kandy=LOC INDEF thirty thousand=SIMIL Malay PL NPST-exist
'There are 30,000 Malays in Kandy.'

The plural marker *pada* is redundant, as the numeral 30,000 already expresses the plurality of *mlaayu*.

Another semantic category that can be expressed redundantly is **semantic class**. For instance, in the Bantu language Kikongo, there are 10 noun classes (Dereau 1955: 17ff). These classes have to be expressed both on the nouns themselves and on adjectives, verbs, demonstratives, etc., as shown in example (7).

Kikongo (Dereau 1955: 39)

(7) N-zo âme ya-mbote i-na.
II.SG-house my II.SG-good II.SG-be
'My house is good.'

Another instance of redundancy is **apposition**: when two (or more) morphosyntactic constituents refer to one semantic object or individual. I speak of **phrasal apposition** when both referencing constituents are lexical. The Tamil sentence in (8) contains an example of phrasal apposition. The constituents *en raŋgan* and the NP *naalu vayaadu* make reference to the same individual. Two constituents are therefore related to one individual: a non-transparent relation.

Tamil (Andronov 2004: 362)

(8) En raŋgan, naalu vaya-adu paiyan, terub-ie alai-v-aan
my Rangan four year-SG.N boy street-LOC run-FUT-3SG.M
'My Rangan, a four-year-old boy, will run about in the street.'

Note that the nature of the relation between the constituents in apposition is not important here. The elements in apposition could be an NP and a resumptive pronoun, an NP and a defining phrase, or any other combination of two referential units. Since this type of apposition is presumably (near-)universal, it does not contribute to the transparency measure. Therefore, I will leave it out of the current study.

However, another type of apposition is not universal and is included. This concerns the phenomenon of cross-reference, where one of the two referencing units is grammatical (e.g. a verbal affix). That term is only applicable when the elements in apposition are able to refer on their own, that is, when using only one of them is also grammatical. For instance, in a pro-drop language, the verbal marker is able to refer on its own, as a verb marked for person and number is grammatical without the pronoun. When an independent pronoun is realized anyway (e.g. in European Portuguese *eu chegue-i* '1SG arrive-1SG'), it is redundant. In FDG, cross-reference is strictly distinguished from agreement (Hengeveld & Mackenzie 2008: 350). The latter term is used only when both referencing units are obligatory, as in French *nous chant-ons* '3PL sing-3PL', where neither the pronoun nor the verbal marker can be left out.

### 4.3 Domain disintegration

A second category of transparency violations is domain disintegration. This involves all structures violating the integrity of a (morphosyntactic or phonological) form. Such violations have repercussions for the way form relates to meaning. For instance, the disintegration of frequent forms can result in their **fusion**, i.e. a two-to-one meaning-to-form relation. Hengeveld (2007) distinguishes between fusion affecting grammatical units (**cumulation**, which is traditionally termed fusional morphology) and fusion affecting lexical stems (**stem alternation**). Referential markers in Gujarati are instances of cumulation: the semantic categories of person, number and tense (three 'meaning units') are expressed in one form, e.g. *ũ* 'PRS.1SG' (Doctor 2004: 41) and *e* 'PRS.3SG' (ibid. p. 54). A very frequent case of cumulation is the combined expression of person and number in pronominal elements. As these categories are so often cumulated, it could be argued that they are actually one and the same semantic unit. However, since languages do exist in which the two are expressed separately, I will analyze this as fusion.

Stem alternation occurs for instance in European Portuguese, where verbal lexical stems are marked for tense and person by suppletion, e.g. the verb *ser* 'to be' has the forms *são* 'they are', *eram* 'they were (IPFV)', *foram* 'they were (PFV)', etc. (J. Lachlan Mackenzie, personal communication). The lexical stem and the grammatical units cannot be formally separated, hence this is opaque. This is an example of a suppletive paradigm, but stem alternation can also be less radical,

for instance in English past tense inflection of strong verbs where only vowels alternate. Note that the cases of fusion described here do not necessarily originate in a process of combining separate forms. The term fusion is used in this paper as indicating that there is a many-to-one relation between meaning and form, not that there has been a diachronic development resulting in the integration of forms.

Another feature in the category of domain disintegration is **discontinuity**: units that belong together pragmatically and semantically are split up in their morphosyntactic realization. **Circumfixes** are discontinuous units by definition. An example comes from Sudanese Arabic, where verbs are marked for person and number by means of vowel patterns in the stem and circumfixes, e.g. *ta-…-ii* '2SG.F.IPFV', *ta-…-uu* '2PL.M.IPFV', etc. (Bergman 2002: 23). The semantic units here (person, number, gender and aspect) are realized in two morphosyntactic units.

**Infixation** can introduce discontinuity in other units. For instance, in English, it is possible to insert certain morphemes in the middle of lexemes, e.g. *bloody* in *abso-bloody-lutely*. Semantically speaking, there is one unit *absolutely*, but at the morphosyntactic level this unit is split up into two separate parts. The result is a one-to-two relation between semantics and morphosyntax.

The **raising** of an argument from an embedded clause into its matrix clause results in discontinuity as well. This is possible in European Portuguese, as illustrated in (9).

European Portuguese (J. Lachlan Mackenzie, p.c.)

(9) O João parece estar doente.
DEF João seems be ill
'João seems to be ill.'

The argument *O João* is morphosyntactically separated from its semantically related predicate, *estar doente*. In other words: the syntactic configuration (João as argument of 'to seem') does not match the semantic configuration (João as argument of 'to be ill').

Discontinuity can furthermore arise as a consequence of **extraposition**, that is: dislocation of a part of a constituent, for instance when a heavy relative clause is moved to the right of a sentence while its head remains in its original position. This is attested in English, cf. example (10).

(10) [The athlete] will be invited [that has won most prizes].

Again, the semantic configuration does not match the syntactic realization in a one-to-one fashion.

Yet another non-transparent feature in the category of domain disintegration is **non-parallel alignment** between different levels of organization. Alignment is the way in which units are grouped together. In a transparent language, groupings

of units at different levels are expected to be parallel. Opacity arises when boundaries at the pragmatic and semantic levels do not coincide with morphosyntactic and phonological groupings. An example of such non-parallel alignment is the English sentence *you'd think*. Phonologically speaking, *you* and *'d* are grouped together, but semantically or pragmatically, they do not form a group. Clitics often attach to semantically unrelated hosts, which makes them a common source of non-parallel alignment (cf. Cysouw 2005).

### 4.4 Form-based form

The third type of non-transparency that I distinguish is form-based form. This category deals with all pragmatically and semantically empty material, in other words: all formal elements and operations that do not have pragmatic or semantic counterparts. Such formal units are there for a morphosyntactic or phonological reason, but do not have meaning.

A first non-transparent feature in this category is **agreement**: the obligatory copying of morphosyntactic information from one unit to another. This can occur at the phrasal level, as in European Portuguese *ess-a mulher timid-a* 'that-F woman(F) shy-F', and in the clausal domain, as in English *he walk-s*, where 3rd person singular is obligatorily copied from the pronoun to the verb. Clausal agreement is of course very similar to cross-reference as defined above — in fact, there are few linguists who would distinguish the two. As explained above, the crucial difference lies in the 'obligatoriness' of the independent element: if it is optional (like in pro-drop languages), I speak of cross-reference, but if it is obligatory, I speak of agreement. The reason for this is that agreement is analyzed in FDG as a purely morphosyntactic copying operation: an obligatory rule without a semantic or pragmatic trigger. When the independent unit is optional, the verb marker cannot be the result of copying.[16] Note furthermore that in a clausal agreement situation, both units are obligatory and cannot occur in isolation — they have no referential power of themselves.

Another form-based form feature is the presence of **expletives**; words like *it* in *it is raining*. *It* does not refer to anything in the world; it is empty.[17] However,

**16.** Unless one assumes that there can be copying from a covert element, as assumed by generative grammarians (and by Hengeveld 2012). The idea is that in some languages, there is agreement between the covert element 'pro' (Hengeveld speaks of an activated participant in the context) and the verbal marker. This possibility will not be pursued further here.

**17.** Apart from these nominal expletives, languages have verbal expletives: copulas are in many languages obligatory because of grammatical rules (e.g. to bear tense), but do not refer to anything. For reasons of time and space, I will restrict myself to nominal expletives.

without *it*, the sentence is ungrammatical — English requires the subject slot to be filled. There is hence a morphosyntactic reason, not a pragmatic or semantic one, to use *it*. A language that does not use expletives is SLM, illustrated in (11).

Sri Lanka Malay — (Nordhoff 2009: 504)
(11) Arà-uujang.
NPST-rain
'It is raining.'

Travis (1984) found an implicational hierarchy for nominal expletive elements, such that if a language uses no expletives in constructions with weather predicates, it does not have them at all. Weather predicates will therefore be used in this study as a litmus test for the presence of expletives.

Another form-based form phenomenon is **grammatical gender**: a classification system of nouns that is not semantically conditioned. An example is the gender system of European Portuguese. Nouns in that language are either masculine or feminine. For animates, the gender is assigned semantically, but gender is not predictable for inanimates (J. Lachlan Mackenzie, p.c.). Hence, Portuguese inanimate nouns have a formal feature that does not have a pragmatic or semantic counterpart. Under this header I will group all nominal classification systems that are non-semantic. This includes systems with partially or completely unpredictable gender assignment, as well as systems that have gender assignment on the basis of morphophonemic information, e.g. when the stem-final phoneme determines inflection. The latter is usually called declension.

A **sequence of tenses rule** (also called tense copying or *consecutio temporum*) is opaque as well. This is a rule requiring the tense of a verb in an embedded clause to adapt to the tense of the main verb. The embedded tense form is then not semantically motivated. English has a sequence of tenses rule at work in the sentence *He said that Munir was ill*. The past tense form of the embedded verb *was* is not semantically based — the sentence can be used when Munir is still ill at the time of the current utterance. The tense form is the result of a morphosyntactic rule, not of a pragmatic or semantic unit. Example (12) illustrates the absence of tense copying in DIP: the past tense of the main clause is not copied to the tense of the embedded clause.

Diu Indo-Portuguese (Cardoso 2009: 135)
(12) Yo sab-iŋ ki el tə fal-a Liza mem.
1SG know-PST COMP 3SG IPFV.NPST say-INF L. EMPH
'I knew that he would say (lit.: is saying) "Liza" only.'

Yet another instance of form-based form is **influence of morphosyntactic weight on word order**. This involves the shifting of morphosyntactically heavy

constituents to the right, and light units to the left — regardless of ordering principles that the language usually obeys. Heavy NP-shift is a well-known example, but other constituents such as verb phrases, clauses or other (parts of) constituents can undergo shifting as well. The opacity lies in the fact that morphosyntactic information determines ordering — word order is transparent only when fully determined by pragmatic and semantic considerations. In example (13) from SLM, the heavy complement clause is moved to post-verbal position, while complements usually occur pre-verbally in SLM.

Sri Lanka Malay (Nordhoff 2009: 739)

(13) Se=ppe oorang thuuva pada anà-biilang
1SG=POSS man old PL PST-say
[kithang pada Malaysia=dering anà-dhaathang katha].
1PL PL Malaysia=ABL PST-come QUOT
'My elders said that we had come from Malaysia.'

Another non-transparent feature is the **influence of morphosyntactic complexity on function marking**. Transparency requires that only pragmatic and semantic functions determine a process like case marking. So in a fully transparent language, all units should be marked identically, regardless of their complexity. But for instance in English, some pronouns are inflected for case while nouns and more complex NPs are not. Morphosyntactic complexity thus influences a morphosyntactic process, overruling pragmatic and semantic information.

It follows that transparent languages will make use of phrase-marking rather than head-marking. After all, if we for instance have a phrase [x y z], where x is the head, it is transparent to mark the entire phrase ([x y z]-f) instead of the head only ([x-f y z]). In the latter scenario, the syntactic scope of function marker f would be x, while semantically speaking, the function scopes over the entire phrase. There would then be a discrepancy between syntax and semantics.

In this study, bound morphemes that scope over words only will be named head-markers, while units like clitics and adpositions that scope over phrases will be named phrase-markers. From the discussion above it follows that the use of clitics rather than affixes is a transparent property of a language. There is no consensus on what a clitic exactly is, in morphosyntactic terms at least. Several criteria have been proposed, of which I will use one as decisive: I will analyze a bound morpheme as a clitic when it is not selective as to its host (it can attach to units of different morphosyntactic classes and of different degrees of complexity) and can hence mark entire phrases. This is in opposition to affixes, which can only

mark the head of a phrase.[18] The non-selectiveness and blindness for morphosyntactic information makes a clitic a transparent unit. In SLM, for example, dative case is marked by means of the clitic *=nang*. The clitic can be used on all kinds of hosts: (pro)nouns, adjectives, clauses, etc. (Nordhoff 2009: 335). Because *=nang* is blind to the morphosyntactic nature of the marked unit, its syntactic and semantic scope always run parallel. A typical affix would only be able to take syntactic scope over a head, while scoping semantically over an entire phrase. A second criterion for clitichood that I will use if necessary is the unit's ability to occur in different positions. For instance, a typical affix like the plural marker *-s* in English can only attach at the right side of a noun, while the person marking verbal clitics in French can attach to the left side (in declarative clauses) and to the right side (in interrogative clauses) of their host.

Another non-transparent phenomenon is that of **syntactic functions**. Predicates have arguments, and these argument have functions; i.e. they fulfill specific roles in the situation or event denoted by the predicate. If pragmatic roles (e.g. Topic/Focus) are relevant for the expression of arguments, there is a one-to-one relation between pragmatics and morphosyntax, which is transparent. If semantic roles (e.g. Actor/Undergoer) determine alignment, there is a transparent relation between semantics and morphosyntax. But when neither pragmatic nor semantic roles are expressed overtly in morphosyntax, there is a *syntactic* function that does not have a 'meaningful' counterpart. I reserve the terms Subject, Object and Indirect Object for such syntactic functions and will not use those terms for semantic or pragmatic roles. SLM shows semantically based alignment (cf. Nordhoff 2011b): semantic roles of arguments are always relevant for their morphosyntactic expression. This is shown in example (14). If the arguments of the different intransitive clauses in (14a–c) would all be marked identically, we would have evidence for morphosyntactic 'neutralization' (i.e. ignoring) of the different semantic functions of the intransitive arguments. This is however not the case: the argument in (14a) is zero-marked, the one in (14b) is marked by accusative case, while in (14c) we find a dative-marker. The semantics of the predicates determines the marking on the argument.

Sri Lanka Malay (Nordhoff 2011b: 100)

(14) a. Dee=Ø su-thiidor baava=ka.
3SG=NOM PST-sleep down=LOC
'He slept downstairs.'

18. The distinction I make between affixes and clitics is gradual, as 'selectiveness' is not a binary notion. This possibly results in classification problems; for instance, if some bound morpheme attaches to simple nouns and verbs, but nothing else, is it a clitic or an affix? For each language, I will try to make out at least what the dominant strategy is (phrase-marking or head-marking).

b. Titanic kappal=yang su-thìnggalam.
Titanic ship=ACC PST-sink
'The ship Titanic sank.'
c. Go=dang karang bannyak thàràsìggar.
1SG.FAM=DAT now very sick
'I am very sick now.'

(15) a. Thora is breaking the vases.
b. The vases are breaking.

Compare this to English: the Actor in (15a) (*Thora*) behaves identically (morpho-syntactically) to the Undergoer in (15b) (*the vases*) — both trigger agreement on the verb and stand in pre-verbal position. The semantic roles Actor and Undergoer are neutralized — one cannot distinguish them on the basis of formal marking. This means that there is evidence for at least a syntactic function Subject in English.

Evidence for a syntactic function Object is formed by the dative shift construction. This is found in English as well, cf. example (16).

(16) a. She insulted him.
b. She gave a present to him.
c. She gave him a present.

In (16a), the Undergoer is in accusative case. In (16b), the Undergoer is not overtly marked, but the Recipient is marked by 'to+ACC'. Now, in (16c), the semantic role Recipient is marked identically as the Undergoer in (16a) — by accusative case, without the preposition. The semantic roles of Recipient and Undergoer are neutralized as to their overt marking. There is a syntactic function Object. Note that if I find a Subject in a language, I will not look further for an Object, as the presence of one syntactic function will qualify the language as non-transparent in this respect.

Yet other form-based form features are alternations caused by morphological, morphophonemic or phonological information. An example is external sandhi in Ambonese Malay: a word-final consonant can be adapted to an adjacent consonant, e.g. *seng b'isa* 'NEG possible' is pronounced [sem b'isa] (Minde 1997: 56). This is place assimilation; similar processes are assimilation of manner of articulation, insertion or deletion of phonemes, tone assimilation, nasalization, etc. These processes all involve changes in phonological form triggered by morphological, morphophonemic or phonological information and are hence opaque. A common source for alternations is the creation of phoneme combinations that are articulatory difficult or morphophonemically forbidden in the language.

## 5. Transparency in creoles: An empirical study

### 5.1 Methodology

To test the hypothesis that contact languages are relatively transparent, four contact languages were tested on their transparency and compared to their respective sub- and superstrate languages. To do this, the list of features introduced in Section 4 was used, repeated in Table 2. As said before, the list is not exhaustive. Yet it includes many features from different areas of grammar (pragmatics, semantics, morphosyntax, and phonology) and is thus able to measure the transparency of languages to a considerable extent. If a language illustrates many of the listed features, it is considered relatively non-transparent, and vice versa. As said in Section 4.1, it is at this moment not possible to assign a value on a scale for each feature. Therefore, a language is said to possess a certain feature as soon as some evidence for its presence is found, even if the feature is present only optionally or

**Table 2.** List of non-transparent phenomena

| Violation type | Non-transparent phenomenon | Subfeature |
|---|---|---|
| Redundancy | Multiple expression of semantic information | Negative concord |
| | | Plural concord |
| | | Semantic class |
| | Apposition | Cross-reference |
| Domain disintegration | Fusion | Cumulation |
| | | Stem alternation |
| | Discontinuity | Circumfixes |
| | | Infixes |
| | | Raising |
| | | Extraposition |
| | Non-parallel alignment | |
| Form-based form | Agreement | |
| | Expletive elements | |
| | Grammatical gender | |
| | Sequence of tenses rule | |
| | Influence of weight on word order | |
| | Influence of complexity on function marking | |
| | Syntactic functions | |
| | Alternations caused by morphological/morphophonemic/phonological information | |

marginally. This might appear over-strict, but this strictness is necessary to guarantee the methodological soundness of the study.

The four creoles studied are Nubi, Pichi, Sri Lanka Malay, and Diu Indo-Portuguese. These are selected because of their diverse linguistic backgrounds; their sources are typologically, genetically, and geographically diverse. As the creole status of Sri Lanka Malay is subject to debate, I will use the term 'contact language' henceforth for the languages, to avoid controversy.

The contact languages will be compared to their source languages. Two problems have come up in the selection of source languages. Firstly, it is impossible to study the source languages as they were spoken at the time of their influence on the contact languages — there are no grammars of these particular languages as spoken during the colonial period. I therefore have had to rely on grammars of these languages as they are spoken nowadays, which are of course different from the languages spoken at the time of contact. Each feature that is attested in this study could have arisen *after* the relevant contact situation. Even though this may somewhat decrease the validity of the results, it does not render them meaningless, as it is unlikely that a language has become completely different from what it was a few hundred years ago.

A second problem is the uncertainty that exists about the exact source languages of the contact languages under consideration. It is, for instance, subject to debate which groups and which languages contributed to Krio (and hence to Pichi). In the end, the choice for selecting some sources over others was made on the basis of different factors, primarily on the basis of theoretical considerations. I selected the languages involved in the account that is in my eyes most plausible. A second reason for including a source language in this study was more practical; in some cases, only one or two of the possible source languages are described well enough to allow for a thorough analysis. The motivations behind the inclusion of particular languages are given in the relevant language analyses. In sum, the sample of languages consists of:

**Table 3.** Languages included in the study

| Contact language and source languages | Language family | References used |
|---|---|---|
| Nubi | | Owens 1990; Wellens 2003; Luffin 2004, 2005; Khamis & Owens 2007; Kihm 2011 |
| Sudanese Arabic | Afro-Asiatic: Semitic | Trimingham 1946; Persson et al. 1979; Bergman 2002; Dickins 2010 |
| Bari | Nilo-Saharan: Nilotic | Spagnolo 1933; Owens 1990; Yokwe 2001 |

**Table 3.** (*continued*)

| Contact language and source languages | Language family | References used |
|---|---|---|
| Pichi | | Smith 2002; Kouwenberg & LaCharité 2004; Yakpo 2009 |
| English | Indo-European: Germanic | Zwicky & Pullum 1983; Kallel 2007 |
| Fongbe | Niger-Congo: Kwa | Lefebvre & Brousseau 2002; Smith 2002; Zeijlstra 2007 |
| Kikongo | Niger-Congo: Bantu | Dereau 1955; Smith 2002 |
| Sri Lanka Malay | | Ansaldo & Nordhoff 2009; Nordhoff 2009; Nordhoff 2011a; Nordhoff 2011b |
| Tamil | Dravidian | Asher 1982; Lehmann 1989; Schiffman 1999; Andronov 2004; Steever 2005 |
| Sinhala | Indo-European: Indo-Aryan | Gair 1970; Gair and Paolillo 1997; Chandralal 2010; Nitz & Nordhoff 2010 |
| Ambonese Malay | Austronesian: Indonesian | Collins 1980; Minde 1997; Paauw 2009; Nordhoff 2011a |
| Diu Indo-Portuguese | | Cardoso 2009 |
| European Portuguese | Indo-European: Romance | Leeuw 1997, Hutchinson & Lloyd 2003, Cardoso 2009 |
| Gujarati | Indo-European: Indo-Aryan | Cardona 1965; Mistry 1997; Doctor 2004; Grosz & Patel 2006; Patel 2007; Cardoso 2009 |

The set of source languages is typologically and geographically diverse, as shown in Table 3. This is necessary to exclude the possibility that structural similarities in the contact languages result from direct inheritance from structurally similar source languages. If we indeed find that these combinations of diverse source languages have resulted in languages that are similar in their degree of transparency, it would be a strong indication that language contact is the explanatory factor.

In the next section, the results of the transparency analyses of all languages will be summarized. For reasons of space, it was not possible to give the complete analysis of each language in this paper. I refer the interested reader to DOI: 10.1075/jpcl.28.2.03leu.additional for more data, argumentation, and examples of all languages.

### 5.2 Results: Contact languages

In Tables 3–5, the occurrence of transparent features in four contact languages is shown per type of opacity. The tables are ordered relative to the transparency

per language (a more transparent language standing more to the left) and the occurrence of features (a more attested feature standing higher, subfeatures being ordered below their main feature). Subfeatures are printed in a smaller typeface. The features are again formulated as non-transparent features (cf. footnote 14). A plus sign means that the *non*-transparent phenomenon *is attested*, hence that the language is opaque with respect to that feature. A minus sign in these tables means that the language lacks the opaque phenomenon and is transparent with respect to that feature. A question mark indicates that I have not been able to find conclusive evidence for the occurrence of that feature in that language. 'N.a.' means that the feature is not applicable for that language. This is used with the feature 'multiple marking of semantic class', for those languages that do not have a semantic classification and therefore cannot possibly mark it redundantly.

**Table 4.** Redundancy in contact languages

| Feature | Nubi | SLM | DIP | Pichi |
|---|---|---|---|---|
| Apposition: Cross-reference | – | – | – | – |
| Multiple expression of semantic information | – | + | + | + |
| Negative concord | – | – | + | + |
| Plural concord | – | + | – | + |
| Semantic class | n.a. | n.a. | – | n.a. |

Table 4 shows that there is redundancy in the contact languages. Multiple expression of semantic information is present in three of the contact languages. However, none of them exhibits cross-reference and semantic class is either not present (Nubi, SLM, Pichi) or not expressed redundantly (DIP).

**Table 5.** Domain disintegration in contact languages

| Feature | SLM | Pichi | DIP | Nubi |
|---|---|---|---|---|
| Fusion | + | + | + | + |
| Cumulation | + | + | + | + |
| Stem alternation | – | – | + | + |
| Discontinuity | – | ? | ? | ? |
| Extraposition | – | ? | ? | ? |
| Raising | – | – | – | ? |
| Circumfixes | – | – | – | – |
| Infixes | – | – | – | – |
| Non-parallel alignment | – | ? | ? | ? |

The presence of redundancy in creoles appears to be a counter example to the expectation that creoles are maximally transparent. But its appearance is not so unexpected if we look at the effect it has: the use of multiple forms to express one meaning actually increases intelligibility.

Table 5 shows that there is domain disintegration in the contact languages as well. Fusion is attested in all four languages, even though they are all predominantly isolating. Cumulation is present in all of them, but stem alternation is not found in SLM and Pichi. It is quite impossible to make any generalizations on discontinuity or non-parallel alignment in contact languages, as too much information is missing here.

**Table 6.** Form-based form in contact languages

| Feature | SLM | Nubi | DIP | Pichi |
|---|---|---|---|---|
| (Morpho-)phonologically triggered alternations | + | + | + | + |
| Influence of weight on word order | + | + | + | + |
| Syntactic functions | – | + | + | + |
| Expletive elements | – | – | – | + |
| Influence of complexity on function marking | – | – | – | – |
| Agreement | – | – | – | – |
| Grammatical gender | – | – | – | – |
| Sequence of tenses rule | – | – | – | – |

In Table 6, we see that also when it comes to form-based form, contact languages are not 100% transparent. (Morpho)-phonologically based alternations are present in all four.[19] This is not so unexpected, since such alternations are to a degree dependent on universal articulatory constraints. Furthermore, all languages exhibit some influence of morphosyntactic weight on word order. Expletives exist in Pichi, and Pichi, DIP, and Nubi have syntactic functions.

However, the studied contact languages exhibit very few form-based form features. There are some in all studied languages, but none of them exhibits agreement, influence of complexity on function marking, *consecutio temporum*, or grammatical gender.[20]

Adding up the results, it turns out that the creoles show a certain amount of redundancy and some domain disintegration as well. However, in the domain of form-based form, the contact languages are quite transparent. While there are

19. Morphologically based alternations seem to be less frequent than phonological ones, but time is lacking to look into this closer. This will remain a topic for future research.

20. This is in line with Trudgill (1999), who finds no creoles with grammatical gender.

differences between the four languages, the similarities with respect to transparency are striking. This similarity cannot be explained by similarities in source languages, as the source languages of the studied creoles are in fact not similar. Hence as explained above, I think the explanation for the relative transparency of creoles should be sought in the effect of language contact. To show that language contact has made the contact languages more transparent than their source languages, a comparison per contact language will be executed in the next section.

## 5.3 Contact languages and parent languages compared

In this section, the data are discussed from a different perspective. The contact languages are compared with their source languages. For each contact language, the features of the source languages can be seen as a feature pool, from which the new language has taken over some but not other features. As the hypothesis given in Section 3 predicts that contact languages will either be as transparent as or more transparent than their sources, I expect that non-transparent features of the source languages will be transparent in contact languages. If all source languages have a plus, but the contact languages has a minus, this is considered strong evidence for the hypothesis. If one or more, but not all of the source languages have a non-transparent property, but the contact language is transparent with respect to that property, this is evidence for the hypothesis as well.[21] However, when a transparent feature is present in at least one of the source languages, but the contact language is non-transparent with respect to that feature, this is taken to be counterevidence to the hypothesis. For some features, lack of information renders it impossible to say whether a (non-)transparent variety was present in the feature pool. These features will be left out of consideration.

### 5.3.1 *Sri Lanka Malay compared to Sinhala, Tamil and Malay*

As can be concluded from Tables 7–9, Sri Lanka Malay has increased its transparency with respect to its source languages. The non-transparent phenomena of cross-reference, stem alternation, influence of complexity on function marking, agreement, *consecutio temporum*, and grammatical gender were all present in at least one source language, but Sri Lanka Malay did not inherit or develop them.

**21.** A reviewer states that a correspondence between a source and a contact language does not mean that the feature was inherited: the feature could also be inherited from another language or have developed on its own. In his opinion then, the absence of an opaque feature where it was present in a source language does not give evidence for the hypothesis. However, I do think it constitutes evidence, as I think the source of the transparent feature in the source language is not relevant — what is relevant is the increase in transparency with respect to the source.

**Table 7.** Redundancy in Sri Lanka Malay and its source languages

| | SLM | Ambon Malay | Tamil | Sinhala |
|---|---|---|---|---|
| Apposition: Cross-reference | – | – | + | – |
| Multiple expression of semantic information | + | – | + | + |
| Negative concord | – | – | – | – |
| Plural concord | + | – | + | + |
| Semantic class | n.a. | n.a. | – | – |

**Table 8.** Domain disintegration in Sri Lanka Malay and its source languages

| | SLM | Ambon Malay | Tamil | Sinhala |
|---|---|---|---|---|
| Fusion | + | + | + | + |
| Cumulation | + | + | + | + |
| Stem alternation | – | – | + | + |
| Discontinuity | – | ? | + | + |
| Extraposition | – | ? | + | + |
| Raising | – | ? | ? | ? |
| Circumfixes | – | – | – | – |
| Infixes | – | – | – | – |
| Non-parallel alignment | – | ? | – | ? |

**Table 9.** Form-based form in Sri Lanka Malay and its source languages

| | SLM | Ambon Malay | Tamil | Sinhala |
|---|---|---|---|---|
| (Morpho-)phonologically triggered alternations | + | + | + | + |
| Influence of weight on word order | + | ? | + | ? |
| Expletive elements | – | – | – | – |
| Syntactic functions | – | + | + | + |
| Influence of complexity on function marking | – | – | + | + |
| Agreement | – | – | – | + |
| Sequence of tenses rule | – | – | + | – |
| Grammatical gender | – | – | + | + |

Moreover, all source languages are opaque with respect to syntactic functions, but SLM is not.

There is a counter example for Sri Lanka Malay: it is more opaque than Ambon Malay with respect to plural concord. This is unexpected, as the hypothesis predicts that Sri Lanka Malay would be at least as transparent as Malay.

Nordhoff (2011b) interestingly states about the high degree of transparency of Sri Lanka Malay: 'This transparency is common in varieties of Malay. It is a retention of a historic feature and not due to language contact' (2011:96). In my opinion, SLM's transparency is both historic (i.e. inherited from Malay) *and* due to language contact; the circumstances have constrained the inheritance, and they have determined which features were inherited and which were not.

### 5.3.2 *Diu Indo-Portuguese compared to Portuguese and Gujarati*

As apparent in Tables 10–12, Diu Indo-Portuguese has increased its transparency compared to its sources in several domains: both Gujarati and Portuguese are opaque with respect to cross-reference, influence of complexity on function marking, agreement, and grammatical gender, but DIP has is transparent with respect to those features. Furthermore, DIP did not inherit plural concord, raising, circumfixes, and sequence of tenses from Portuguese, but is fully transparent regarding those features.

However, Diu Indo-Portuguese shows negative concord while Gujarati does not. DIP apparently inherited this non-transparent feature from Portuguese,

**Table 10.** Redundancy in Diu Indo-Portuguese and its source languages

| | **DIP** | **Gujarati** | **Portuguese** |
|---|---|---|---|
| Apposition: Cross-reference | – | + | + |
| Multiple expression of semantic information | + | – | + |
| Negative concord | + | – | + |
| Plural concord | – | – | + |
| Semantic class | – | n.a. | n.a. |

**Table 11.** Domain disintegration in Diu Indo-Portuguese and its source languages

| | **DIP** | **Gujarati** | **Portuguese** |
|---|---|---|---|
| Fusion | + | + | + |
| Cumulation | + | + | + |
| Stem alternation | + | + | + |
| Discontinuity | ? | ? | + |
| Extraposition | ? | ? | + |
| Raising | – | ? | + |
| Circumfixes | – | – | + |
| Infixes | – | – | – |
| Non-parallel alignment | ? | ? | + |

**Table 12.** Form-based form in Diu Indo-Portuguese and its source languages

| | DIP | Gujarati | Portuguese |
|---|---|---|---|
| (Morpho-)phonologically triggered alternations | + | + | + |
| Influence of weight on word order | + | ? | + |
| Expletive elements | – | – | – |
| Syntactic functions | + | + | + |
| Influence of complexity on function marking | – | + | + |
| Agreement | – | + | + |
| Sequence of tenses rule | – | – | + |
| Grammatical gender | – | + | + |

which is unexpected since the transparent option was also present in the DIP feature pool.

### 5.3.3 *Nubi compared to Bari and Sudanese Arabic*

Nubi has gained in transparency compared to its source languages, as shown in Tables 13–15. It did not take on plural concord or agreement, which are present

**Table 13.** Redundancy in Nubi and its source languages

| | Nubi | Bari | Sudanese Arabic |
|---|---|---|---|
| Apposition: Cross-reference | – | – | ? |
| Multiple expression of semantic information | – | + | + |
| Negative concord | – | ? | – |
| Plural concord | – | + | + |
| Semantic class | n.a. | – | + |

**Table 14.** Domain disintegration in Nubi and its source languages

| | Nubi | Bari | Sudanese Arabic |
|---|---|---|---|
| Fusion | + | + | + |
| Cumulation | + | + | + |
| Stem alternation | + | + | + |
| Discontinuity | ? | ? | + |
| Extraposition | ? | ? | – |
| Raising | ? | ? | + |
| Circumfixes | – | – | + |
| Infixes | – | – | – |
| Non-parallel alignment | ? | ? | ? |

**Table 15.** Form-based form in Nubi and its source languages

| | Nubi | Bari | Sudanese Arabic |
|---|---|---|---|
| (Morpho-)phonologically triggered alternations | + | + | + |
| Influence of weight on word order | + | ? | + |
| Expletive elements | – | ? | – |
| Syntactic functions | + | ? | ? |
| Influence of complexity on function marking | – | – | – |
| Agreement | – | + | + |
| Sequence of tenses rule | – | – | – |
| Grammatical gender | – | – | – |

in both Bari and Sudanese Arabic. The opaque phenomenon of double expression of semantic class, exhibited by Sudanese, is also lost in Nubi. There is no feature where Nubi increased its opacity with respect to its sources.

### 5.3.4 *Pichi compared to English, Fongbe, and Kikongo*

Pichi is more transparent than its source languages, as can be concluded from Tables 16–18. It did not take on the opaque phenomena of cross-reference, double

**Table 16.** Redundancy in Pichi and its source languages

| | Pichi | English | Fongbe | Kikongo |
|---|---|---|---|---|
| Apposition: Cross-reference | – | – | – | + |
| Multiple expression of semantic information | + | + | + | + |
| Negative concord | + | + | – | ? |
| Plural concord | + | + | + | + |
| Semantic class | n.a. | n.a. | n.a. | + |

**Table 17.** Domain disintegration in Pichi and its source languages

| | Pichi | English | Fongbe | Kikongo |
|---|---|---|---|---|
| Fusion | + | + | + | + |
| Cumulation | + | + | + | + |
| Stem alternation | – | + | – | + |
| Discontinuity | ? | + | + | + |
| Extraposition | ? | + | + | ? |
| Raising | – | + | – | ? |
| Circumfixes | – | – | – | + |
| Infixes | – | + | – | – |
| Non-parallel alignment | ? | + | ? | ? |

**Table 18.** Form-based form in Pichi and its source languages

| | Pichi | English | Fongbe | Kikongo |
|---|---|---|---|---|
| (Morpho-)phonologically triggered alternations | + | + | + | + |
| Influence of weight on word order | + | + | ? | ? |
| Expletive elements | – | + | – | ? |
| Syntactic functions | + | + | + | ? |
| Influence of complexity on function marking | – | + | – | + |
| Agreement | – | + | – | – |
| Sequence of tenses rule | – | + | – | ? |
| Grammatical gender | – | – | – | – |

marking of semantic class, stem alternation, circumfixes, infixes, expletives, influence of complexity on function marking, agreement, or sequence of tenses, even though these were all present in one or more source languages. Negative concord is the only counter example: while at least Fongbe is transparent with respect to negative concord, Pichi is not.

## 6. Discussion and conclusions

In Section 3, the hypothesis of this study was formulated: all contact languages are either as transparent or more transparent than their source languages. This hypothesis is corroborated on the basis of the results presented in Section 5.2 and 5.3. All contact languages 'gained' transparency with respect to their source languages: many non-transparent features of the source languages are lost so that the contact languages turn out to have a higher degree of transparency compared to their sources.

Sri Lanka Malay, Diu Indo-Portuguese, and Pichi do not exhibit cross-reference, while this was present in one of their sources. Furthermore, double expression of semantic class has been lost in Nubi and Pichi. But agreement, influence of complexity on function marking, sequence of tenses, and grammatical gender are especially often lost. These features appear in none of the contact languages. This is striking, as they are present in at least six (agreement), seven (influence of complexity on function marking), three (sequence of tenses), and four (grammatical gender) of the ten source languages. This suggests that loss of opacity through language contact is most salient in the domain of form-based form: if a language loses an opaque feature, it will most likely lose a form or operation that has no pragmatic or semantic counterpart.

There are three counter examples to the hypothesis. Negative concord is present in Diu Indo-Portuguese and in Pichi, while one of their respective source languages (Gujarati and Fongbe, respectively) lacks negative concord. Sri Lanka Malay exhibits plural concord, while Ambon Malay does not. In these cases, the contact languages have a non-transparent feature when a transparent option was available as well. Crucially, all counter examples are cases of redundancy — one semantic unit (negation or plurality) is expressed more than once. This can be explained by the fact that redundancy, unlike domain disintegration and form-based form, increases intelligibility. This could very well have been the reason that these features were inherited (or developed language-internally), even though they are opaque.

It turns out that form-based form is the rarest and 'most severe' type of non-transparency, while redundancy is present even in relatively transparent languages. Domain disintegration takes a place in between. On this basis, the implicational hierarchy in Figure 1 can be stipulated.

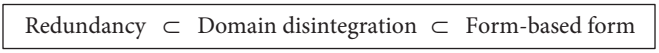


**Figure 1.** Hierarchy of opacity

This hierarchy states that if a language has *some* form-based form feature(s), it will also have some domain disintegration feature(s) and some redundancy feature(s). Note that it need not be the case that *all* redundancy features are present in a language before the first domain disintegration appears.

A sample of four contact languages and ten source languages is rather small to form a basis for claims of universality. Moreover, the sample is not a valid representation of all creoles or contact languages of the world — I have made no attempt to draw a geographically or genetically sound sample. Still, this hierarchy can be seen as a tendency and a starting point for further research. Leufkens (to appear) will provide more evidence for this hierarchy on the basis of a larger sample, as well as refine it by ordering all separate opaque features in a multi-dimensional hierarchy.

Further research on the transparency of many more languages should lead to a better understanding of the coherence of mismatch features. What looks like a random collection of incomparable features at first sight turns out to be explained by the same factors, for instance intelligibility. Addressing these factors will in the end lead us to an answer on the question: why are languages not fully transparent?

*Received: 3/9/2011*
*Revised: 6/4/2012*
*Accepted: 6/11/2012*

## Abbreviations

| | | | |
|---|---|---|---|
| 1, 2, 3 | 1st, 2nd, 3rd person | M | masculine |
| I, II, III, etc. | semantic class | N | neuter |
| ABL | ablative | NEG | negation |
| ACC | accusative | NOM | nominative |
| AOR | aorist | NPST | non-past |
| COMP | complementizer | PASS | passive voice |
| COND | conditional | PFV | perfective |
| DAT | dative | POSS | possessive |
| DEF | definite | PROG | progressive |
| EMPH | emphatic | PRS | present tense |
| F | feminine | PST | past tense |
| FAM | familiar | PL | plural |
| FUT | future | QUOT | quotative |
| GEN | genitive | REL | relativizer |
| INDEF | indefinite | SG | singular |
| INF | infinitive | SIMIL | simulative |
| IPFV | imperfective | SUBJ | subject |
| LOC | locative | | |

## References

Aboh, Enoch. 2009. Competition and selection. That's all! In Enoch Aboh & Norval Smith (eds.), *Complex processes in new languages*, 317–344. Amsterdam: John Benjamins.

Aboh, Enoch & Norval Smith. 2009. Simplicity, simplification, complexity and complexification. Where have the interfaces gone? In Enoch Aboh & Norval Smith (eds.), *Complex processes in new languages*, 1–25. Amsterdam: John Benjamins.

Aboh, Enoch & Umberto Ansaldo. 2009. The role of typology in language creation. In Umberto Ansaldo, Stephen Matthews, & Lisa Lim (eds.), *Deconstructing creole*, 39–67. Amsterdam: John Benjamins.

Aksu-Koç, Ayhan & Dan Slobin. 1985. The acquisition of Turkish. In Dan Slobin (ed.), *The cross-linguistic study of language acquisition*, 839–878. Hillsdale: Lawrence Erlbaum Associates.

Andronov, Mikhail S. 2004. *A reference grammar of the Tamil language*. Munich: LINCOM Europa.

Ansaldo, Umberto & Sebastian Nordhoff. 2009. Complexity and the age of languages. In Aboh & Smith (eds.), *Complex processes in new languages,* 345–363. Amsterdam: John Benjamins.

Arends, Jacques, Pieter Muysken & Norval Smith (eds.). 1995. *Pidgins and creoles: An introduction*. Amsterdam: John Benjamins.

Asher, Ron. 1982. *Tamil* (Lingua Descriptive Studies 7). Amsterdam: North-Holland Publishing Company.

Bakker, Peter. 2004. Phoneme inventories, language contact and grammatical complexity: A critique of Trudgill. *Linguistic Typology* 8(3). 368–375.

Bakker, Peter, Aymeric Daval-Markussen, Mikael Parkvall, & Ingo Plag. 2011. Creoles are typologically distinct from non-creoles. *Journal of Pidgin and Creole Languages* 26(1). 5–42.
Bergman, Elizabeth M. 2002. *Spoken Sudanese Arabic: Grammar, dialogues and glossary*. Springfield: Dunwoody Press.
Blom, Elma, Daniela Polisenska, & Fred Weerman. 2008. Articles, adjectives and age of onset: The acquisition of Dutch grammatical gender. *Second Language Research* 24. 297–331.
Braun, Maria & Ingo Plag. 2003. How transparent is creole morphology? A study of early Sranan word-formation. In Geert Booij & Jaap van Marle (eds.), *Yearbook of morphology 2002*, 81–104. Dordrecht: Kluwer.
Cardona, George. 1965. *A Gujarati reference grammar*. Philadelphia: The University of Pennsylvania Press.
Cardoso, Hugo. 2009. *The Indo-Portuguese language of Diu*. Utrecht: LOT dissertation.
Chandralal, Dileep. 2010. *Sinhala*. Amsterdam: John Benjamins.
Collins, James T. 1980. *Ambonese Malay and creolization theory*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
Cysouw, Michael. 2005. Morphology in the wrong place. A survey of preposed enclitics. In Wolfgang Dressler, Dieter Kastovsky, Oscar Pfeiffer, & Franz Rainer (eds.), *Morphology and its demarcations*, 17–37. Amsterdam: John Benjamins.
Dahl, Östen. 2004. *The growth and maintenance of linguistic complexity*. Amsterdam: John Benjamins.
Dereau, Leon. 1955. *Cours de Kikongo*. Namur: Wesmael-Charlier.
Dickins, James. 2010. Basic sentence structure in Sudanese Arabic. *Journal of Semitic Studies* 55(1). 237–261.
Doctor, Raimond. 2004. *A grammar of Gujarati*. Munich: LINCOM Europa.
Dryer, Matthew S. & Martin Haspelmath (eds.). 2011. *The world atlas of language structures online*. Munich: Max Planck Digital Library. Available at: http://wals.info/. (January, February, and March 2011.)
Gair, James. 1970. *Colloquial Sinhalese clause structures*. The Hague: Mouton.
Gair, James & John Paolillo. 1997. *Sinhala*. Munich: LINCOM Europa.
Grosz, Patrick & Pritty Patel. 2006. Long distance agreement and restructuring predicates in Kutchi Gujarati. http://web.mit.edu/pritty/www/patel_grosz_lda.pdf (22 September, 2011.)
Hengeveld, Kees. 2007. Parts-of-speech systems and morphological types. *ACLC Working Papers* 2. 31–48.
Hengeveld, Kees. 2011. Transparency in functional discourse grammar. *Linguistics in Amsterdam* 4. 1–22.
Hengeveld, Kees. 2012. Referential markers and agreement markers in functional discourse grammar. *Language Sciences* 34(4). 468–479.
Hengeveld, Kees & J. Lachlan Mackenzie. 2008. *Functional discourse grammar: A typologically-based theory of language-structure*. New York: Oxford University Press.
Hutchinson, Amélia P. & Janet Lloyd. 2003. *Portuguese: An essential grammar*. Routledge Essential Grammars. New York: Routledge.
Kallel, Amel. 2007. The loss of negative concord in standard English: Internal factors. *Language Variation and Change* 19. 27–49.
Khamis, Cornelia & Jonathan Owens. 2007. Nubi (Creole Arabic). In John Holm & Peter L. Patrick (eds.), *Comparative creole syntax: Parallel outlines of 18 creole grammars*, 199–216. London: Battlebridge.

Kihm, Alain. 2000. Are creole languages 'perfect' languages? In John McWhorter (ed.), *Language change and language contact*, 163–199. Amsterdam: John Benjamins.
Kihm, Alain. 2011. Pidgin-creoles as a scattered sprachbund: Comparing Kriyol and Nubi. *Journal of Pidgin and Creole Languages* 26(1). 43–88.
Kouwenberg, Sylvia & Darlene LaCharité. 2004. Echoes of Africa: Reduplication in Caribbean Creole and Niger-Congo languages. *Journal of Pidgin and Creole Languages* 19(2). 285–331.
Kouwenberg, Sylvia & Darlene LaCharité. 2011. The typology of Caribbean Creole reduplication. *Journal of Pidgin and Creole Languages* 26(1). 194–218.
Kusters, Wouter. 2003. *Linguistic complexity: The influence of social change on verbal inflection*. Utrecht: LOT dissertation.
Langacker, Ronald. 1977. Syntactic reanalysis. In Charles Li (ed.), *Mechanisms of syntactic change*, 56–139. Austin: University of Texas Press.
Leeuw, Frank van der. 1997. *Clitics: Prosodic studies*. The Hague: HIL dissertations.
Lefebvre, Claire. 2001. On the semantic opacity of creole languages. *Journal of Pidgin and Creole Languages* 16. 321–353.
Lefebvre, Claire & Anne-Marie Brousseau. 2002. *A grammar of Fongbe*. Berlin/NY: Walter de Gruyter.
Lehmann, Thomas. 1989. *A grammar of modern Tamil*. Pondicherry: Pondicherry Institute of Linguistics and Culture.
Leufkens, Sterre. To appear. *Transparency in language. A typological study*. Utrecht: LOT Dissertation.
Luffin, Xavier. 2004. *Kinubi texts*. Munich: LINCOM.
Luffin, Xavier. 2005. *Un créole Arabe: Le Kinubi de Mombasa, Kenya*. Munich: LINCOM.
Lupyan, Gary & Rick Dale. 2010. Language structure is partly determined by social structure. *PLoS ONE* 5(1). 1–10.
McWhorter, John. 1998. Identifying the creole prototype: Vindicating a typological class. *Language* 74. 788–818.
McWhorter, John. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5. 125–166.
Miestamo, Matti. 2006. On the feasibility of complexity metrics. In Krista Kerge & Maria-Maren Sepper (eds.), *Conference proceedings of the Annual Finnish and Estonian Conference of Linguistics, May 6–7, 2004*, 11–26. Tallin: TLÜ.
Minde, Don van. 1997. *Malayu Ambong: Phonology, morphology, syntax*. Leiden: Research School CNWS PhD thesis.
Mistry, P.J. 1997. Gujarati phonology. In Alan S. Kaye (ed.), *Phonologies of Asia and Africa*, 653–674. Winona Lake: Eisenbrauns.
Nitz, Eike & Sebastian Nordhoff. 2010. Subtractive plural morphology in Sinhala. http://www.eva.mpg.de/lingua/staff/nordhoff/pdf/PluralInSinhala_4.pdf (5 September, 2011.)
Nordhoff, Sebastian. 2009. *A grammar of upcountry Sri Lanka Malay*. Utrecht: LOT dissertation.
Nordhoff, Sebastian. 2011a. The genesis of Sri Lanka Malay as a multi-layered process. http://www.eva.mpg.de/lingua/staff/nordhoff/pdf/The_development_of_SLM_as_a_multi-layered_process.pdf (26 October, 2011.)
Nordhoff, Sebastian. 2011b. Transparency in Sri Lanka Malay. *Linguistics in Amsterdam* 4. 96–110.
Omar, Margaret. 1973. *The acquisition of Egyptian Arabic as a native language*. The Hague: Mouton.
Owens, Jonathan. 1990. East African Nubi: Bioprogram vs. inheritance. *Diachronica* 7(2). 217–250.

Paauw, Scott H. 2009. *The Malay contact varieties of Eastern Indonesia: A typological comparison*. Buffalo: State University of New York PhD dissertation.
Parkvall, Mikael. 2008. The simplicity of creoles in a cross-linguistic perspective. In Matti Miestamo, Kaius Sinnemäki, & Fred Karlsson (eds.), *Language complexity. Typology, contact, change*, 265–285. Amsterdam: John Benjamins.
Patel, Pritty. 2007. *Split ergativity and the nature of agreement*. London: University College MA thesis. http://web.mit.edu/pritty/www/patel2007.pdf (22 September, 2011.)
Persson, Andrew, Janet Persson, & Ahmad Hussein. 1979. *Sudanese colloquial Arabic for beginners*. Horsleys Green: SIL.
Schiffman, Harold F. 1999. *Spoken Tamil*. Cambridge: Cambridge University Press.
Seuren, Pieter & Herman Wekker. 1986. Semantic transparency as a factor in creole genesis. In Pieter Muysken (ed.), *Substrata versus universals in creole genesis*, 57–70. Amsterdam: John Benjamins.
Slobin, Dan. 1977. Language change in childhood and history. In John Macnamara (ed.), *Language learning and language thought*, 185–214. New York: Academic Press.
Smith, Norval. 2002. The history of Surinamese creoles II. Origin and differentiation. In Eithne Carlin & Jacques Arends (eds.), *Atlas of the languages of Suriname*, 131–151. Leiden: KITLV Press.
Snow, Catherine, Norval Smith, & Marian Hoefnagel-Höhle. 1980. The acquisition of some Dutch morphological rules. *Journal of Child Language* 7. 539–553.
Spagnolo, Lorenzo M. 1933. *Bari grammar*. Verona: Missioni Africane.
Steever, Sanford. 2005. *The Tamil auxiliary verb system*. Routledge Asian Linguistics Series. New York: Routledge.
Thomason, Sarah. 2001. *Language contact*. Edinburgh: Edinburgh University Press.
Travis, Lisa. 1984. *Parameters and effects of word order variation*. Cambridge: MIT dissertation.
Trimingham, John Spencer. 1946. *Sudan colloquial Arabic*. London: Oxford University Press.
Trudgill, Peter. 1999. Language contact and the function of linguistic gender. *Poznán Studies in Contemporary Linguistics* 35. 133–152.
Trudgill, Peter. 2004. Linguistic and social typology: The Austronesian migrations and phoneme inventories. *Linguistic Typology* 8. 305–320.
Wellens, Ineke. 2003. *An Arabic creole in Africa. The Nubi language of Uganda*. Nijmegen: Catholic University of Nijmegen dissertation.
Yakpo, Kofi. 2009. *A grammar of Pichi*. Nijmegen: Radboud University of Nijmegen dissertation.
Yokwe, Eluzai. 2001. *The tonal grammar of Bari*. Ann Arbor: UMI dissertation.
Zeijlstra, Hedde. 2007. Negation in natural language: On the form and meaning of negative elements. *Language and Linguistics Compass* 1(5). 498–518.
Zwicky, Arnold M. & Geoffrey K. Pullum. 1983. Cliticization vs. inflection: English n't. *Language* 59. 502–513.

## *Author's address*

Sterre Leufkens
Universiteit van Amsterdam
Departement Taal — en letterkunde
Spuistraat 210
1012 VT Amsterdam

s.c.leufkens@uva.nl