

Complexity between Explicitness and Economy— Overt vs. Hidden Complexity

0. Basic idea of the paper

- (i) There is a special type of complexity based on pragmatics, i.e. hidden complexity (§ 2)
- (ii) Hidden complexity is significantly higher in East and mainland Southeast Asia (EMSEA) than elsewhere. Even Creoles generally show a lower degree of hidden complexity. The contact languages determine the degree of hidden complexity. (§ 3)
- (iii) The high degree of hidden complexity in EMSEA languages reflects an economy-oriented process of maturation (cf. Dahl 2004 and his explicitness-oriented approach). This process is based on at least the following three factors that determine frequency: (i) what is already there, (ii) morphology that does not express inflectional categories and (iii) language contact. (§ 4)

1. Economy vs. explicitness and overt vs. hidden complexity as two types of complexity

1.1. Approaches to complexity generally discussed in linguistics

- *Hidden simplicity*

The complex structures we find in the grammars of the world's languages can be reduced to a small set of recursive processes.

Formal Linguistics: Hauser, Chomsky & Fitch (2002)

Functional Linguistics: Givón & Shibatani (2009)

For Hauser, Chomsky & Fitch (2002), recursion is the only property of language that is specifically human / that is part of Faculty of Language in a narrow sense (FLN):

- *Overt complexity*

This approach is surface-oriented: It looks at grammatical categories as they are overtly marked (McWhorter 2001, 2005; Dahl 2004, Miestamo et al. 2008, Sinnemäki 2011)

To compare complexity across languages, McWhorter proposes four diagnostics of grammatical complexity that are based on the following intuition:

(1) Complexity (McWhorter 2005: 45):

[A]n area of grammar is more complex than the same area in another grammar to the extent that it encompasses more overt distinctions and/or rules than another grammar.

The **four diagnostics** in some detail (McWhorter 2005: 45 – 46):

- (i) A phonemic inventory is more complex to the extent that it has more marked members
- (ii) A syntax is more complex than another to the extent that it requires the processing of more rules, such as asymmetries between matrix and subordinate clauses.
- (iii) A grammar is more complex than another to the extent that it gives overt and grammaticalized expression to more fine-grained semantic and/or pragmatic distinctions than another
- (iv) Inflectional morphology renders a grammar more complex than another one in most cases

There is a third type complexity => hidden complexity.

2. Hidden complexity

2.1. The historical background of hidden complexity

Languages differ essentially in what they must convey and not in what they may convey. (Jakobson 1992 [1959]: 149)

Taking Sapir's (1921: 82, 90) example:

(1) English: *The farmer kills the duckling.*

=> *English*: Various obligatory categories: number, definiteness, tense, person/number agreement of the subject with the verb.
Chinese: Almost no obligatory categories.

Sapir's (1921) transformation of obligatoriness as it is reflected in (1) into Chinese is still determined by the general idea that arguments have to be expressed. Thus, the brackets in (2) are mine [W.B]. As we will see, radical pro-drop is widespread in EMSEA.

(2) (*Farmer*) *kill* (*duck*).

2.2. Why are there differences in terms of obligatory categories? — The articulatory bottleneck and explicitness vs. economy

The informational incompleteness of language: Morphosyntactic structures and their properties can never fully express the meaning they have in a concrete speech situation. They need pragmatic enrichment from context.

The reason for this is the “**articulatory bottleneck**” (2000: 6, 27-30). Human speech is characterized by its extremely slow transmission rate—processes like prearticulation, parsing and comprehension can produce and assess much more information in a much shorter time:

“inference is cheap, articulation expensive” (Levinson 2000: 29).

The articulatory bottleneck has its impact on the grammatical systems of individual languages:

Grammatical structures can be seen as the result of the competition between expensive articulation or **explicitness** and cheap inference or **economy**: => **Competing motivations**:

- von der Gabelentz (1891: 251): *Deutlichkeitstrieb* vs. *Bequemlichkeitstrieb*
- Haiman (1983): iconic vs. economic motivations
- Optimality Theory: faithfulness vs. markedness constraints

The competition between explicitness and economy creates two types of complexity:

Explicitness => **Overt complexity**:
 The structure of the language simply forces the speaker to explicitly encode certain grammatical categories even if they could easily be inferred from context.

Economy: => **Hidden complexity** (Bisang 2009, forth):
 The structure of the language does not force the speaker to use a certain grammatical category if it can be pragmatically inferred from context.

2.3. The manifestations of hidden complexity

Hidden complexity refers to **unexpressed grammatical categories which need to be inferred by the hearer**. The unexpressed grammatical categories **can be expressed non-obligatorily** if needed in the language.

Grammatical systems of individual languages allow different degrees of hidden complexity.

Apart from the necessity of inferring unexpressed/hidden categories, hidden complexity triggers additional more specific effects:

- Impossibility of determining unique reference in cases such as coreference and reference tracking.
- If a non-obligatory marker represents a certain construction its absence creates a surface structure that stands for more than one construction. The construction intended by the speaker has to be inferred again from linguistic and non-linguistic context.

Thus, seemingly simple sequences of words may represent a considerable number of different constructions.

The phenomena to be discussed are:

- Zero arguments (*zero pronouns, null subjects/null objects, (radical) pro-drop*; Lack of obligatory category) (cf. § 3.1)
- Relative-clause formation and multiple coreference options through zero marking (more than one coreference option) (cf. § 3.2)
- Clause combining (lack of overt construction marking, the same surface structure represents more than one construction) (cf. 3.3)

The languages to be discussed are:

- The languages with high hidden complexity (HC) properties: EMSEA languages: Mandarin Chinese (Sinitic)
- Creole languages, which mostly show lower HC complexity, depending on contact:
 - Atlantic Creoles:
 - Angolar: Portuguese-based, spoken on São Tomé, an island in the Gulf of Guinea, Africa
 - Berbice Dutch: Dutch-based, spoken in Guyana, South America
 - Haitian Creole or Kreyòl Ayisyen: French-based, spoken in Haiti.
 - Non-Atlantic Creoles:
 - Mauritian Creole: French-based, spoken on Mauritius Island (SW Indian Ocean)
 - Tok Pisin: English-based, a variety of Melanesian Pidgin spoken in Papua New Guinea
 - Zamboangueño: Spanish-based, a variety of Chabacano or Philippine Creole Spanish, spoken in and around Zamboanga City on the southern tip of Mindanao island.
- Important contact languages for many Atlantic Creoles: Kwa languages: Fongbe (Niger-Congo: Kwa), Yoruba (Niger-Congo: Benue-Congo: Defoid)

3. Hidden complexity in Chinese, Kwa languages and Creoles

3.1. Zero arguments

The classic approach to pro-drop (Rizzi 1986): Correlation between rich agreement morphology for person and number and the option of dropping the subject argument. In his approach, the *pro* element must be *formally* licensed (by a head of C, I, V, P ...) and *by content* (through rich agreement). What this account cannot explain is **radical pro-drop** (also

rampant pro-drop, discourse pro-drop), in which person/number features cannot be retrieved from overt marking:

- (4) a. 你昨天看了電影嗎？ — 看了。
 Nǐ zuótiān kàn-le diànyǐng_i ma? — \emptyset_i kàn-le \emptyset_i .
 you yesterday see-PFV film Q see-PFV
 ‘Did you see a film yesterday?’ — ‘[I] saw [one].’

Four solutions to the problem of radical pro-drop are discussed in the literature (cf. Ackema et al. 2006, Neeleman & Szendrői 2007):

- (i) Licensing of poor agreement
- (ii) Topic drop
- (iii) Blocking by determiners
- (iv) Requirement of agglutinating morphology on the pronouns

None of these accounts works, either for reasons of theoretical inconsistencies or for empirical counterexamples.

3.1.1. Zero arguments in Chinese

Chinese is a **radical pro-drop language** (cf. example (4)). Zero marking occurs even in cases of subject change (unmarked change of subject!):

- (5) Chinese: Lack of overt arguments (from Lu Xun, *Fēigōng* ‘Opposing aggression’):
 子夏的徒弟公孙高来找墨子，已经好几回了，总是不在家，见不着。大约是第四或者第五回罢，这才恰巧在门口遇见，...

Zǐxià de túdì Gōngsūn Gāo_i lái zhǎo Mò-zǐ_j, yǐjīng hǎojiǔ
 Zixia POSS disciple Gongsun Gao come seek Mo-Master already several
 huí le, \emptyset_j zǒngshì bú zài jiā, \emptyset_i jiàn bu zháo \emptyset_j . Dàyuē shì
 times PF never NEG be.at home see NEG reach about be
 dì-sì huòzhě dì-wǔ huí ba, zhè cái qiàqiǎo zài ménkǒu
 fourth or fifth time EXCL this finally by.chance at doorway
 \emptyset_i yùjiàn \emptyset_j , ...
 meet

‘Gongsun Gao_i, a disciple of Zixia, was looking for master Mo_j for several times and [he_j] was never at home, so [he_i] was unable to meet [him_j]. It was at about the fourth or the fifth time that [he_i] met [him_j] in the doorway ...’

Radical pro-drop is not a new phenomenon in Chinese. It holds at least since classical Chinese (5th – 3rd centuries BC):

- (6) Classical Chinese (Zhuangzi 4.1), lack of overt arguments:
 顏回見仲尼，請行。曰奚之？曰將之衛。曰奚為焉？

Yán Huí_i jiàn Zhòng Ní₂, \emptyset_1 qǐng \emptyset_1 xíng. \emptyset_2 Yuē: \emptyset xī zhī? \emptyset_2 Yuē: \emptyset jiāng
 Yan Hui see Confucius ask leave say where go say FUT
 zhī Wèi. \emptyset_2 Yuē: \emptyset xī wéi yān?
 go Wei say what do there

‘Yan Hui saw Confucius and asked [him for the permission to] leave. [Confucius] said: “Where do [you] go?” [Hui] replied: „[I] will go to Wei”. [Confucius] asked: “What do [you] do there?”’

3.1.2. Zero arguments in some West African Kwa languages: Fongbe

In most West African languages, there are at least two sets of pronouns, free/independent pronouns and dependent/clitic pronouns (sometimes divided into subcategories). This is also the case in Fongbe:

	Personal Pronouns	Clitics [+nominative]	Clitics [-nominative]
1. SG	<i>nyè</i>	<i>ùn</i>	<i>mì</i>
2. SG	<i>hwè</i>	<i>à</i>	<i>wè</i>
3. SG	<i>é(yè)</i>	<i>é</i>	<i>è</i>
1. PL/2.PL	<i>mí</i>	<i>mí</i>	
3. PL	<i>Yé</i>	<i>yé</i>	

Table 1: Pronominal system of Fongbe (Lefebvre & Brousseau 2002: 63)

In Fongbe, simple independent clauses with a transitive verb have an obligatory subject and an obligatory object. There are, however, a few verbs called “verbs licensing expletive subjects” by Lefebvre & Brousseau (2002: 246, 276-277) with optional subjects. In example (7), the expletive is optional, in (8) it is obligatory:

(7) Lefebvre & Brousseau (2002: 276):

(É) *cí qɔ̃ Kɔ́kú jè àzɔ̃n.*
 3.SG seem COMP Koku be.sick
 ‘It seems that Koku is sick.’

(8) Lefebvre & Brousseau (2002: 277):

É *vè-wí nú Kɔ́kú ní yì.*
 3.SG be-difficult COMP Koku SUBORD leave
 ‘It is difficult that Koku leaves.’

With simple transitive verbs, the object slot is not allowed to be empty as long as the object is not extracted to the focus position or is a *wh*-word (Lefebvre & Brousseau 2002: 247f).

3.1.3. Zero arguments in Creole languages

A large number of Creoles makes a distinction between strong and weak pronouns in which the weak forms behave like syntactic clitics (cf. the similarities with Kwa). Therefore, many Creoles can be analysed as **null subject languages** but not as radical pro-drop languages.

Atlantic Creoles:

Radical pro-drop seems to be rather limited in Atlantic Creoles (also cf. Neeleman & Szendrői 2007: Haitian, Jamaican and Papiamentu are no radical pro-drop languages):

Haitian: There are constructions with no overt subject marking. Example (9) illustrates the raising verb *genle* ‘seem’, which cannot take an expletive subject. Other constructions without expletives are existential predicates and weather predicates. Elsewhere, overt arguments are necessary (10) (cf. Déprez 1994, against DeGraff 1993).

(9) Haitian Creole (DeGraff 1993: 72):

- a. *ø genle Jak damou.* b. **li genle Jak damou.*
 seem Jack be.in.love 3.SG seem Jack be.in.love
 ‘Jack seems to be in love.’

(10) Haitian Creole (DeGraff 1993: 72):

- | | |
|--------------------------------------------------|------------------------------------------------|
| a. <i>Li pati.</i>
3.SG leave
'S/He left.' | b. <i>*\emptysetpati</i>
leave |
|--------------------------------------------------|------------------------------------------------|

Non-Atlantic:

Tok Pisin: Various grammars state that the subject must be marked overtly in simple independent clauses (e.g. Verhaar 1995):

(11) Tok Pisin (Verhaar 1995: 35):

- Em i kam insait gen.*
3.SG PM come on again
'He came in once ahain.'

A look at texts reveals **that there are null referential subjects**. Meyerhoff (2000: 134) found null subjects for third person singular in 39% of the clauses analyzed in her sample.

Zamboangueño does not need to have an overt subject (since it is a VSO language, the \emptyset -markers are put after the verb):

(12) Zamboangueño (Lipski & Santoro 2007: 376):

- Andá \emptyset alyì na réyno.*
go there to kingdom
'He goes there to the kingdom.'

(13) Zamboangueño (Lipski & Santoro 2007: 376):

- Kwándo sale \emptyset afwéra ya murí \emptyset .*
when leave outside PST die
'When he went outside, he died.'

Lipski & Santoro (2007: 376, fn. 2) mention a number of zero-pronouns in various contexts. Since Philippine-type Austronesian languages are generally pro-drop, the pro-drop properties of Zamboangueño seem to be due to substrate influence.

3.1.4. Conclusion on zero arguments

Chinese	Radical pro-drop/zero arguments
Fongbe	No zero argument marking
Creoles	Saramaccan: No radical pro-drop, null subjects Haitian Creole: No radical pro-drop, overt marking of arguments in most cases Angolar: No radical pro-drop, but zero marking occurs in discourse Tok Pisin: Zero marking occurs in discourse Mauritian Creole: zero argument is possible but with certain restrictions Zamboangueño: wide-spread zero marking

Table 2: Summary on zero argument marking

- Zero arguments are much further developed in Chinese than in West African languages and in most Creoles.
- The degree to which zero arguments and hidden complexity are possible is related to the substrate and superstrate languages involved.
 - It is comparatively low in languages with West African background (Saramaccan, Haitian, Angolar)
 - It is higher and allows radical pro-drop/zero arguments if the substrate language allows it (Zamboangueño) or after further development of the Creole.

3.2. Relative clauses

3.2.1. Relative clauses in Chinese

Chinese relative clauses always end with the attributive marker *de* and precede their head nouns. The syntactic function and the semantic role of the head noun within the relative clause can be inferred from the fact that arguments of intransitive and monotransitive verbs must be zero-marked. Relative constructions consisting merely of a transitive verb are ambiguous. The head noun may be the subject (14a) or the object (14b) of the relative clause:

(14) 找的人還沒有回來。

[zhǎo de] rén hái méiyǒu huí-lái.

Look.for REL man/people still NEG:PST return-come

a. Object coreference:

‘The people [(we) were looking for] haven’t come back yet.’

b. Subject coreference:

‘The people [who looked (for us)] haven’t come back yet.’

With certain head nouns and in certain contexts (cf. Ning 1993, Wang 2003), the head noun can also be interpreted as a genitive (15a), as a locative (15b) or as an instrumental (15c):

(15) a. Possessor coreference:

那個 [頭髮很長的] 學生

nà ge [tóufa hěn cháng de] xuésheng

that CL hair very long REL student

‘the student whose hairs are very long’ (Wang 2003)

b. Locative coreference:

[他吃春卷的] 饭馆

[tā chī chūnjuǎn de] fànguǎn

s/he eat spring.roll REL restaurant

‘the restaurant where s/he ate spring rolls’

c. Instrumental coreference:

[我写信的] 毛笔

[wǒ xiě xìn de] máobǐ

I write letter REL brush/pencil

‘the pencil I write a letter with’

3.2.2. Relative clauses in West Africa: Fongbe

Even though relative-clause formation is quite different in West African languages, there is a clear-cut distinction between subject coreference, object coreference and non-argument coreference. Grammar does not allow instances in which coreference must be pragmatically inferred.

In Fongbe, the head noun “is linked to a position within the relative clause through the lexical nominal operator, *dé*” (Lefebvre & Brousseau 2002: 161). This operator attracts lexical material. In the case of relative clauses, it attracts resumptive pronouns by moving them from their initial position in the relative clause to the position immediately after *dé*. This process generates surface structures which clearly differ for each type of coreference:

(16) a. Subject coreference (Lefebvre & Brousseau 2002: 161):

sínû [dɛ-é wá] ɔ

man OP-3.SG.SBJ come DEF

‘the man who came’

b. Object coreference (Lefebvre & Brousseau 2002: 161):

àsón_i [dɛ-è_i Kòkú d̀̀̀ ɔ_i]

crab OP-3.SG.OBJ Koku eat DEF

‘the crab that Koku ate’

- c. Locative coreference (Lefebvre & Brousseau 2002: 161):
xàsùn_i [dɛ́-é-mɛ́ Kòkú sɔ̀wû dɔ́ ø_i] ɔ́.
 basket OP-3.SG.OBJ-in Koku take cat put DEF
 ‘the basket in which Koku put the cat’

3.2.3. Relative clauses in Creole languages

There is always a morphosyntactic distinction between argument and non-argument coreference in the Creole languages I have looked at. In two Atlantic Creoles, there is also an obligatory subject/object asymmetry: Haitian, Angolar. In the other languages, the subject vs. object distinction is facultative. Some examples:

Creole language	Relative-clause formation/distinctions in coreference marking
Haitian	Argument/non-argument distinction, Subject/object asymmetry (relativizer <i>ki</i> vs. \emptyset)
Angolar	Argument/non-argument distinction, Subject/object asymmetry (relativizer <i>ki</i> vs. <i>ma</i>)
Berbice Dutch	Argument/non-argument distinction, facultative distinction of subject vs. object coreference
Tok Pisin	Argument/non-argument distinction, facultative distinction of subject vs. object coreference
Zamboangueno	Argument/non-argument distinction, facultative distinction of subject vs. object coreference

Table 3: Summary of hidden complexity in Creole relative clause formation

Angolar (Maurer 1995: 55-58): The marker *ki* is used with subject coreference (17), while *ma* occurs with non-subject coreference (18) and (19):

- (17) Angolar: Relative clause with subject coreference (Maurer 1995: 55):

ome si [ki ba tamba]
 man DEF REL go catch.fish
 ‘the man who left to catch fish.’

- (18) Angolar: Relative clause with object coreference (Maurer 1995: 55):

ome si [ma m bê]
 man DEF REL 1.SG see
 ‘the man I saw’

- (19) Angolar: Relative clause with non-argument coreference (Maurer 1995: 56):

[[ome si]_i [ma n ga taba ra ø_i]]
 man DEF REL 1.SG TA work give
 ‘the man for whom I work’

In **Zamboangueno**, relative-clause formation seems to be less complex than in many other Creoles if one looks at the description provided by Lipski & Santoro (2007: 383-384). Relative clauses are usually (but not compulsorily) introduced by the relativizer *ke* or *kyén*. There is a difference between arguments and non-arguments. Non-argument coreference (22) implies pied-piping (preposition/case marker *kon* ‘with’ plus relativizer).

- (20) Zamboangueno: Relative clause with subject coreference (Lipski & Santoro 2007: 383):

el mana hénte_i [kyén ya man tunúk ø_i na gargánta]
 DEF PL people REL PST DRV be.prick.by.thorn in throat
 ‘people who have gotten fish spines caught in their throat’

(21) Zamboangueno: Relative clause with object coreference (Lipski & Santoro 2007: 383):

El hombre_i, [ke ya man enkontrá tu ϕ_i], mi hermano.
 DEF man RELPST DRV meet 2.SG my brother
 ‘The man whom you met is my brother.’

(22) Zamboangueno: Relative clause with non-argument coreference

(Lipski & Santoro 2007: 383):

El persona [kon-kyen ta kombersá tu] byen bwéno gayót
 DEF person with-REL PROG talk 2:SG very good EMPH
 ‘The person you are talking to is very nice indeed.’

3.2.4. Conclusion on relative clauses

- Chinese shows the highest degree of hidden complexity, Yoruba and Fongbe the lowest.
- All creoles make a distinction between [\pm argument]-coreference. Thus, they show less hidden complexity than Chinese and more hidden complexity than the African languages.
- Haitian and Angolar also show subject/object asymmetry. They both have African substrate languages which tend to make that distinction.
- Creoles in which the subject/object asymmetry is facultative often show constructional variation in relative-clause formation (presence or absence of a resumptive pronoun). This type of variation is characteristic of many contact situations.

3.3. Clause combining

3.3.1. Clause combining in Chinese

In Chinese, clauses can be juxtaposed without marking the syntactic and semantic relation between them. Each interpretation is associated with different constructions which must be analysed differently and which have different syntactic properties (I agree with Paul 2008 that examples such as (23) stand for different syntactic structures, *pace* Li & Thompson 1973).

(23) Li & Thompson (1981: 595):

我買票進去。

Wǒ mǎi piào jìn-qù.

I buy ticket enter-go

a. **Purpose:** ‘I bought a ticket to go in.’

b. **Consecutive action:** ‘I bought a ticket and went in.’

3.3.2. Clause combining in West Africa: The case of Yoruba¹

In Yoruba, the unmarked juxtaposition of two clauses is extremely context dependent. It is possible in a highly specific context with clause serialization in the focus position. Example (24), which is similar to example (23), is a possible answer in the following context:

There is a soccer game and there are various options to get access to it. Some people come in with the governor, while others have to buy a ticket for entering. If somebody asks *Báwo lo se wọlé?* ‘How did you get in’, (24) is one possible answer, in which the actions of ‘buying ticket’ and ‘enter’ have become closely related through context:

(24) *Mo ra tíkẹ̀tì wọlé ní.*

I buy ticket enter FOC

‘I entered by buying a ticket.’

¹ I owe the examples on Yoruba to Remi Sonaiya, with whom I am working together on Yoruba. I would like to thank her for her help.

Otherwise, purpose and consecutive action must be expressed by two different, overtly marked constructions (cf. (25a) and (25b), respectively), i.e., different semantic relations require different constructions:

(25) a. Purpose:

Mo ra tɪkɛ̀tɛ̀tì láti wɔ́lé.
I buy ticket PURP enter
'I bought a ticket to go in.'

b. Consecutive action:

Mo ra tɪkɛ̀tɛ̀tì, mo sì wɔ́lé.
I buy ticket I and enter
'I bought a ticket and went in.'

3.3.3. Clause combining in Creole languages

Creoles generally use different constructions. The omission of the relation marker (conjunction) does not produce ambiguous surface structures. In most Creole languages, there is a relatively general marker which corresponds to English *and* and various other constructions with other conjunctions. Examples from Berbice Dutch:

(26) Berbice Dutch: *an* 'and' for consecutive action (Kouwenberg 2007: 49):

o mu-tɛ mu waʃi an eni ku-tɛ o an mu-tɛ metɛ o.
3.SG go-PF PURP wash and 3.PL catch-PF 3.SG and go-PF with 3.SG
'He went to bath [at the river side] and they kidnapped him [lit. took him and went away with him].'

(27) Berbice Dutch: *an* 'and' for simultaneous action (Kouwenberg 2007: 49):

di kɛnɛ masi wa tan an kiki-a ju moi.
DEF person must PST stand and see-IPFV 2.SG good
'The person must have been standing [there] and spying you out.'

(28) Berbice Dutch: purpose (and other functions) in general (Kouwenberg 1994: 313):

o wa mu-tɛ fi kopo gut fə fi selfju.
3.SG PST go-PF PURP buy thing for 3.POSS self
'She had gone to buy things for herself.'

The unmarked juxtaposition of two clauses is relatively frequent in conditionals. Since conditionals are used with different TA markers, there are contexts in which the omission of the conditional marker (*afi*, *aso*, *as*, *if*) triggers different interpretations. In (29), the combination of past plus perfect generates hypothetical meaning if *aso* 'if' is there. Without *aso*, the first clause must be interpreted as anterior tense 'He had gotten it here. It killed him':

(29) Berbice Dutch: Conditionals (Kouwenberg 1994: 115):

aso wa krik-ɪt-o hiri, o wa bato doto.
if PST get-PF-3.SG here 3.SG PST kill-:PF:3.SG dead
'If he had gotten it here, it would have killed him.'

3.3.4. Conclusion on clause combining

- The extent to which the omission of conjunctions is possible in Creoles/contact languages and in Yoruba is much smaller than in Chinese.
- Creoles are similar to their lexifier languages (English, French, etc.) in various ways and have thus about the same degree of overt complexity.

3.4. Summary on hidden complexity

To what extent does the grammar of a language leave options for pragmatic inference of grammatical information, i.e., to what extent does it show hidden complexity (HC)?

	Chinese	West Africa: Yoruba, Fongbe	Creoles
Zero arguments	Radical pro-drop/ high degree of HC	No pro-drop	Less inference than in Chinese
Relative clause	Inference of subj vs. object coreference	Inference not necessary	Less inference than in Chinese
Clause combining	Non-marking provides options for inference/high degree of HC	Non-marking has specific meaning	Non-marking provides limited options for inference

Table 4: Summary

In general:

The degree to which hidden complexity in zero arguments, relative clauses and clause combining is possible in Creoles depends to a considerable extent (apart from later internal changes in Creoles) on the hidden-complexity properties of the superstrate and substrate languages involved.

4. Factors that favour the high degree of hidden complexity in EMSEA

4.1. The relevance of frequency

The criterion for the successful diffusion of a feature within a speech community is **frequency**, i.e., the frequency with which a certain feature of the feature-pool occurs in the utterances of a speech community. If one takes the S-curve model of language change (Wang & Cheng 1970, Bailey 1973, Chambers and Trudgill 1998: 162-164), the frequency that is crucial for triggering change is situated between 20 – 30% (cf. the initial stasis, rapid rise, tailing off).

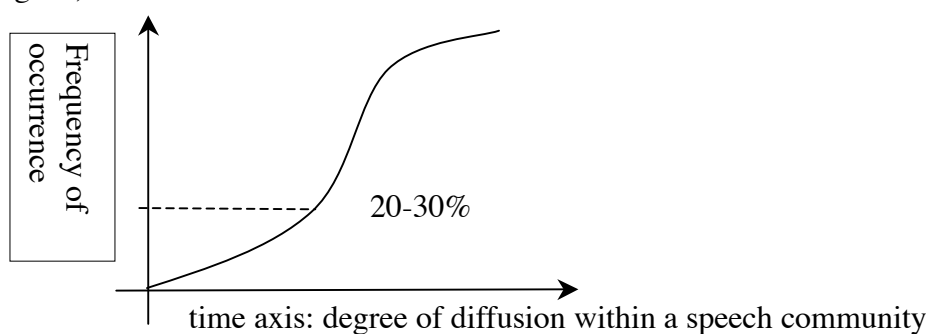


Figure 1: S-curve

This also applies for the change from radical pro-drop to non-pro-drop and for an obligatory explicit marking of coreference relation in relative clauses.² As soon as the frequency of 20 –

² By overt expression of features, I mean any type of overt morphosyntactic expression: (i) syntactic expression by a lexical noun or a pronoun, (ii) morphological expression by a clitic or an affix. In both cases, the features that are relevant for a subject or an object are clearly retrievable from what is explicitly said, i.e., pragmatic inference is not necessary.

30% in the population of relevant utterances is reached, change to non-radical pro-drop or explicit coreference distinctions can be expected within a short period of time.

However, frequency as such is not the complete explanation. There are various factors that determine frequency. If we want to understand, why there are not more languages or groups of languages in which economy is as prominent as in EMSEA languages we need to look at these factors and to provide a **potential scenario** of how they interacted to create the specific situation as we find it in EMSEA languages today.

=> To understand what prevents economy from developing beyond a certain degree (and thus favours explicitness and overt complexity) means to understand what factors can keep frequency low.

The factors that are relevant for EMSEA languages are:

- (i) What is already there:
Language development is always based on certain grammatical structures that already exist and that trigger certain frequency patterns. The properties of these structures have an impact on the extent to which hidden complexity can be developed at later stages.
- (ii) Morphology that does not express inflectional categories:
Inflectional morphology is defined by obligatoriness (cf. Bybee 1985) and thus enhances frequency and explicitness against economy and hidden complexity.
- (iii) Sociolinguistic aspects and contact-induced convergence:
Sociolinguistic factors have their own impact on frequency and on the selection of morphosyntactic expression formats.

4.2. What is already there and the expression of inflectional morphology

4.2.1. Chinese (Sinitic)

That there was morphology in the earlier periods of Old Chinese (OC: 11th – 3rd centuries BC) is uncontroversial. The following list of the most important reconstructed affixes with their functions is based on Sagart (1999):

- *s-: denominal verbs, causative, directive (acts/states directed toward external condition or other persons), inchoatives (?)
- *m-: Controlled actions by volitional actors: (a) volitionality, (b) agentive nouns, (c) small animals
- *k-: actions/objects “that are well-delimited in time and space, and hence usually concrete and countable” (Sagart 1999: 107). With verbs: transient/attempted action, repeated action, continued action.
- *t-: (a) stative verbs, (b) involuntary physiological actions, (c) some other intransitives, (d) some nouns
- *-r-: General idea of intensification: (a) iteratives, (b) action taking place in two or more locations, (c) nouns denoting double/multiple object (e.g. eyeballs), (d) intense quality.
- *-s: (a) denominal nouns, (b) exodirectional/endodirectional verbs (e.g. hear vs. ask)

An example:

(30) Functions of OC *s- [MC stands for Middle Chinese between about 600 and 900 AD]:

a. verbs derived out of object-denoting lexemes (Sagart 1999: 71):

帚	zhǒu	MC	*tsy ^h uwX	OC	* ^b tu?	‘a broom’
掃	sǎo	MC	*sawX	OC	* ^a s-tu?	‘to broom’

b. nouns derived out of verbs (Sagart 1999: 73):

拽 <i>yì</i>	MC * <i>yet/yejH</i>	OC * <i>^blat(-s)</i>	‘to pull’
鞅 <i>xiè</i>	MC * <i>sjet</i>	OC * <i>^bs-hlat</i>	‘leading-string’

c. causatives (Sagart 1999: 70):

食 <i>shí</i>	MC * <i>zyik</i>	OC * <i>^bm-lík</i>	‘to eat’
飼 <i>sì</i>	MC * <i>ziH</i>	OC * <i>^bs-lík-s</i>	‘to feed (tr.)’

d. directives (Sagart 1999: 71):

易 <i>yì</i>	MC * <i>yek</i>	OC * <i>^blek</i>	‘to exchange’
賜 <i>cì</i>	MC * <i>sjeH</i>	OC * <i>^bs-hlek-s</i>	‘to give’

e. inchoatives (Sagart 1999: 72):

悟 <i>wù</i>	MC * <i>nguH</i>	OC * <i>^aŋa-s</i>	‘to be awake, aware’
蘇 <i>sū</i>	MC * <i>su</i>	OC * <i>^as-ŋa</i>	‘to come back to life; to wake up’

Inflectional morphology in Old Chinese is minimal. Some categories listed under (31) may be interpreted as inflectional markers in some languages:

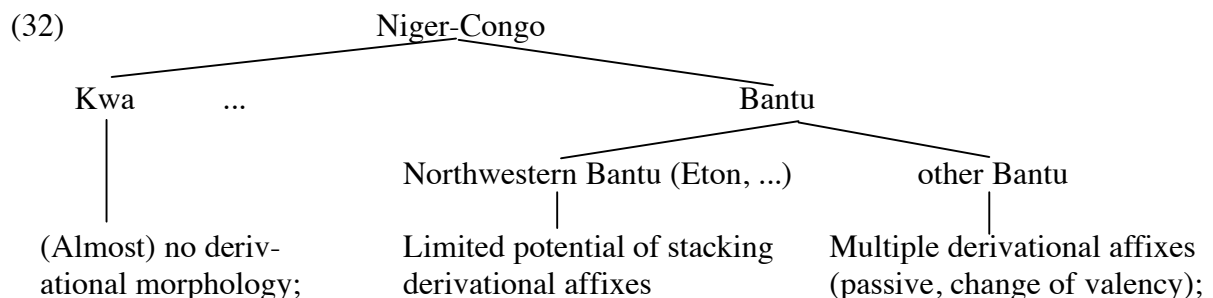
- (31) • Aspect and phase (temporal distribution of event): **s*-for inchoatives, **r*- iteratives, **k*- repeated action
- Causativity: **s*- for denominal causative
 - Valency/voice (e.g. active, passive): **t*- for intransitive verbs and stative verbs
 - Personation (action on self vs. other): **s*- directives (maybe), **-s*: exodirectional/ endodirectional verbs (e.g. hear vs. ask)

but there are at least two reasons why these markers do not reach the frequency they could reach in principle to become inflectional markers in Old Chinese:

- (i) Their meaning is still too specific/concrete, i.e., it interacts with the meaning of the stem (cf. relevance in terms of Bybee 1985).
- (ii) One and the same grammatical concept can be expressed by more than one marker, i.e., Old Chinese morphology is far from having a one-to-one correlation between meaning and form.

4.1.2. Niger-Congo

A preliminary remark on the genealogical structure of the Niger-Congo family (simplified!):



The Kwa languages of the Niger-Congo family (32) are characterized by their loss of verbal morphology (including agreement). In spite of this, these languages are far from being radical pro-drop.

The reason for that:

In general, Proto-Niger-Congo and Proto-Bantu seemed to have considerably more morphology than Chinese and the marking of person on the verb seems to be uncontroversial even if there is a controversial discussion concerning the extent of morphological fusion.

- Güldemann (2007, 2010): Proto-Bantu and Proto-Niger-Congo lacked inflectional verb prefixes.
- Hyman (2004, 2010): Multiple suffixation and prefixation must have existed even in languages which have lost them.
- Nurse (2007: 254) on Bantu:

It is likely that the language which gave rise to today's Bantu languages spoken in eastern Nigeria or western Cameroon, maybe early in the third millennium B.C, had an analytic verb structure at an early point in its development, as probably had the several related languages spoken round it. During the following centuries, cliticisation of the several pre-stem components moved it toward a synthetic structure. On the basis of the admittedly incomplete evidence examined here, I have to admit with some frustration that I cannot be sure whether late common Proto-Bantu was fully synthetic, or whether fully synthetic structures only developed later, after the proto-language had splintered. I am inclined to think it was fully synthetic, just because it is the most economical solution, but cannot be entirely sure.

Even without deciding which of the reconstructions is correct, what remains is the presence of inflectional marking of categories like person. If this is the case, there is a very long continuity of overt presence of arguments. Maybe Hyman's (2010: 24) assumption of a tendency "to cycle back and forth, grammaticalizing full words as inflectional proclitics and prefixes, losing them, and creating them once more" over a period of some 10.000 – 12.000 years is the most plausible scenario.

My hypothesis is that the high frequency with which person/number features occurred in discourse in Niger-Congo was maintained in the syntax of the Kwa languages by creating obligatory syntactic expressions once the obligatory morphological expression has gotten lost.

4.2. Sociolinguistic aspects and contact-induced convergence

- *The frequency of grammatical categories in discourse in EMSEA language families:*
Sinitic strongly supported the low frequency of inflectional grammatical features since Old Chinese. Mon-Khmer morphology had similar properties. As far as one can conclude from relatively recent data from Thai (13th century) and Hmong-Mien, the situation was again similar.
- *The frequency of grammatical categories at the time of contact:*
The languages of EMSEA are characterized by their long and highly complex contact history (Enfield 2003, 2005, Bisang 1996, 2006). At the time when the speakers got in contact, their languages had already reached a high degree of hidden complexity as one can assume from what was already there. Thus, the grammars of the languages involved in contact situations must have allowed speakers to be more economic than in most contact situations that led to grammars of Creole languages.

- *How contact further reduced reminiscences of low hidden complexity:*

Once the high degree of hidden complexity was reached in the individual language families involved in EMSEA, language contact added its own contribution to economy and even eliminated reminiscences of older stages with less hidden complexity. This can be seen from modern Khmer. Its morphology is comparatively well developed if compared to other Mon-Khmer languages. The question of why it did not develop more systematically seems to be due to syntax-based techniques that have been developed in other EMSEA languages:

The formation of agent nouns by means of morphology in Khmer:

- (33) The infix *-m-* in Khmer:

<i>sò:m</i> ‘ask, ask a favour’	=>	<i>smò:m</i> ‘beggar’
<i>cam</i> ‘wait for, guard, keep’	=>	<i>chmam</i> ‘guard, n.’
<i>cù:əŋ</i> ‘do business’	=>	<i>chmù:əŋ</i> ‘business-man’

- (34) The more productive syntax-based method of word formation with *nèək* ‘person’ in the word-initial head position:

nèək-daə(r) [person-walk] ‘pedestrian’
nèək-taəŋ [person-compose/write] ‘author, composer, writer’
nèək-chlò:p [person-go stealthily to watch someone] ‘spy, snoop’.

=> The development of a grammatical system with low-frequency grammatical features is partly due to the input situation (no morphological marking of grammatical categories when the languages got in contact first) and it is further enhanced by contact situations that point into the same direction.

4.3. Maturation

McWhorter (2001, 2005) and the short history of Creoles as a reason for low overt complexity:

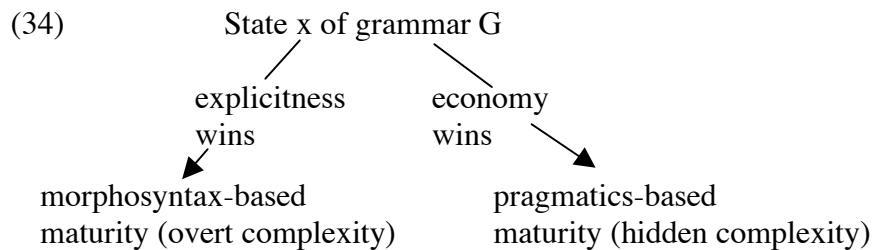
The general conclusion was that in older grammars, millennia of grammaticalization and reanalysis have given overt expression to often quite arbitrary slices of semantic space, the result being a great deal of baroque accretion which, while compatible with Universal Grammar, is incidental to it, as well as to even nuanced human expression. In not having existed for long enough a time for drift to encrust them in this manner to any great extent, creoles are unique in reflecting the innate component of the human language capacity more closely than other languages do. (McWhorter 2001: 126)

Grammars of individual languages develop through time. Dahl (2004: 103-105) looks at grammars of individual languages from the perspective of states and how a certain state *x* developed from a former state *y* (cf. the reconstruction of the moves in a chess game from its actual state *x*). The criteria for determining evolutionary complexity are based on overt marking:

- (33) The accumulation of material in a grammar *G* of a language that did not exist at an earlier stage *G'* of that language.

This is the perspective of explicitness. **From the perspective of economy and pragmatic inference, there is an other type of maturation that leads to hidden complexity.** Thus, at any state *x* of the grammar of a language, there is a bifurcation that either leads to overt

complexity or hidden complexity, if one takes the competing motivations of explicitness vs. economy seriously.



East and mainland Southeast Asian languages show a consistent development to pragmatics-based maturity. This is manifested in the higher relevance of pragmatics in general and the three phenomena analyzed in § 3.

In each of these cases, certain grammatical indicators which are obligatory in other languages are not necessarily expressed and thus call for pragmatic inference of the following grammatical information:

- (i) Radical pro-drop: **Person/Number of subject or object argument**
- (ii) Relative clauses: **Semantic/Syntactic role of head in RC**
- (iii) Clause combining: **Syntactic status of the clauses involved, semantic relation between them**

7. Conclusions

- Complexity has not only an overt side but also a hidden side, which is based on economy and the pragmatic inference of grammatical categories that are not obligatorily expressed in a language.
- EMSEA languages are characterized by a particularly high degree of hidden complexity in the case of (i) radical pro-drop, multiple coreference interpretations in relative clauses and (iii) clause combining.
- Even in Creoles as instances of extreme language contact the degree of hidden complexity is lower, at least in most of them. Creole grammars seem to represent the limits of what can be reduced/omitted in an utterance from the perspective of the structural potential provided by the contact languages involved. Thus, they are just normal languages that evolved under certain social situations (pace McWhorter 2001, 2005).
- The high degree of hidden complexity in EMSEA is due to the following factors that block frequency:
 - Historical preconditions: What is already there
 - Morphology that does not express inflectional categories
 - Language contact that further enhances the diffusion of constructions that favour hidden complexity
- The situation in EMSEA languages is due to an economy-oriented process of maturation.

References

- Ackema, P., Brandt, P., Schoorlemmer, M. and Weerman, F. (eds.). 2006. *Arguments and agreement*. Oxford: Oxford University Press.
- Bailey, C-J N. 1973. *Variation and Linguistic Theory*. Washington, DC: Center for Applied Linguistics.
- Bisang, Walter. 1996. Areal typology and grammaticalization: Processes of grammaticalization based on nouns and verbs in East and mainland South East Asian languages. *Studies in Language*, 20.3, 519-597.
- Bisang, Walter. 2006. 'South East Asia as a Linguistic Area', in: Brown, Keith. (ed.) *Encyclopedia of Language and Linguistics*, Vol. 11, 587-595. Oxford: Elsevier.
- Bisang, Walter. 2008. 'Grammaticalization and the areal factor: The perspective of East and mainland Southeast Asian languages', in: López-Couso, María José & Elena Seoane (eds.), *Rethinking Grammaticalization. New Perspectives*, 15-35. Amsterdam & Philadelphia: John Benjamins.
- Bisang, Walter. 2009. 'On the evolution of complexity — sometimes less is more in East and mainland Southeast Asia', in: Sampson, Geoffrey, David Gil and Peter Trudgill (eds.), *Language Complexity as an Evolving Variable*, 34 – 49. Oxford: Oxford University Press.
- Bisang, Walter. 2010. 'Grammaticalization in Chinese—a construction based account', in: Traugott, Elisabeth C. and Graeme Trousdale (eds), *Gradience, Gradualness, and Grammaticalization*, 245-277. Amsterdam and Philadelphia: Benjamins.
- Bisang, Walter. 2011. Grammaticalization and typology, in: Narrog, Heiko and Heine, Bernd (eds), *Handbook of grammaticalization*, 105-117. Oxford: Oxford University Press.
- Bisang, Walter. forth. Language contact between geographic and mental space. In: Auer, P., M. Hilpert, A. Stukenbrock and B. Szendrői, *Linguistic Perspectives on Space: Geography, Interaction, and Cognition*. Berlin: Mouton de Gruyter (Series Linguae et Litterae)
- Bybee, Joan L. 1985. *Morphology: A Study of the Relation between Meaning and Form*. Amsterdam and Philadelphia: Benjamins.
- Chambers, J. K. & Peter Trudgill. 1998. *Dialectology*. 2nd edition. Cambridge: Cambridge University Press.
- Dahl, Östen. 2004. *The Growth and Maintenance of Linguistic Complexity*. Amsterdam/Philadelphia: Benjamins.
- DeGraff, Michel Frederic. 1993. Is Haitian a Pro-Drop language? In: Francis Byrne and John Holm (eds.), *Atlantic Meets Pacific: A Global View of Pidginization and Creolization*, 71–90. Amsterdam and Philadelphia: Benjamins.
- Déprez, Viviane. 1994. 'Haitian Creole: A Pro-Drop language?', in: *Journal of Pidgin and Creole Languages* 9, 1–24.
- Enfield, Nicholas J. 2003. *Linguistic epidemiology. Semantics and grammar of language contact in mainland Southeast Asia*. London & New York: RoutledgeCurzon.
- Enfield, Nicholas J. 2005. Areal linguistics and Mainland Southeast Asia. *Annual Review of Anthropology* 34, 181-206.
- Gabelentz, G. von der. 1891. *Die Sprachwissenschaft, ihre Aufgaben, Methoden und bisherigen Ergebnisse*. Leipzig: T.O. Weigel Nachfolger. Republished 1972 by Tübingen: Narr.
- Givón, Talmy and Shibatani, Masayoshi. 2009. *Syntactic Complexity. Diachrony, Acquisition, Neuro-Cognition, Evolution*. Amsterdam and Philadelphia: John Benjamins.
- Güldemann, Tom. 2007. 'Bantu in its macro-areal context of Africa and implications for the early typology of Bantu and Niger-Congo.' Paper read at the International Conference on Bantu Languages, Gothenburg: Oct 4–6, 2007.
- Güldemann, Tom. 2010. 'Proto-Bantu and Proto-Niger-Congo: Macro-areal typology and linguistic reconstruction', To appear in Christa König & Osamu Hieda (eds), *International Symposium of the Center of Corpus-Based Linguistics and Language Education (CbLLE)*. Amsterdam & Philadelphia: John Benjamins.
- Haiman, J. 1983. Iconic and economic motivation. *Language* 59: 781–819.

- Hauser, M. D., Chomsky, N., and Fitch, T. W. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298: 1569–1579.
- Hyman, Larry. 2004. 'How to become a Kwa verb', in: *Journal of West African Languages* 30, 69-88.
- Hyman, Larry. 2010. 'The Macro-Sudan Belt and Niger-Congo reconstruction', September 04, 2011: linguistics.berkeley.edu/~hyman/Hyman_Niger_Congo_Dynamics.pdf
- Jakobson, Roman. 1992 [1959]. On linguistic aspects of translation. In: Rainer Schulte and John Biguenet (eds.), *Theories of Translation: An Anthology of Essays from Dryden to Derrida*, 144-151. Chicago and London: The University of Chicago Press. Originally in: Reuben A. Brower (ed.), *On Translation*. Cambridge: Harvard University Press.
- Kouwenberg, Silvia. 1994. *A Grammar of Berbice Dutch Creole*. Berlin: Mouton de Gruyter.
- Lefebvre, Claire & Brousseau, Anne-Marie. 2002. *A Grammar of Fongbe*. Berlin: Mouton de Gruyter.
- Levinson, S. C. 2000. *Presumptive Meanings. The Theory of Generalized Conversational Implicatures*. Cambridge, Mass.: The MIT Press.
- Li, Charles N. & Thompson, Sandra A. 1973. 'Serial verb constructions in Mandarin Chinese: subordination or coordination?', in: *Chicago Linguistic Society* 9, 96-103.
- Li, Charles N. and Thompson, Sandra A. 1981. *Mandarin Chinese. A functional reference grammar*, Berkeley, Los Angeles and London: University of California Press.
- Lipski, John M. and Maurizio Santoro. 2007. 'Zamboangueno Creole Spanish', in: Holm, John and Peter L. Patrick (eds.), *Comparative Creole Syntax. Parallel Outline of 18 Creole Grammars*, 373-396. London: Battlebridge.
- Maurer, Philippe. 1995. *L'angolar. Un créole afro-portugais parlé à São tomé*. Hamburg: Buske.
- Meyerhoff, Miriam. 2000. *Constraints on Null Subjects in Bislama (Vanuatu)*. Canberra: Pacific Linguistics.
- McWhorter, J. H. 2001. 'The world's simplest grammars are creole grammars.', in: *Linguistic Typology* 5, 125–166.
- McWhorter, J. H. 2005. *Defining Creole*. Oxford: Oxford University Press.
- Matti Miestamo, Kaius Sinnemäki and Fred Karlsson. eds. 2008. *Language Complexity: Typology, Contact, Change*. Amsterdam: Benjamins.
- Neeleman, Ad and Kriszta Szendrői. 2007. 'Radical pro drop and the morphology of pronouns', in: *Linguistic Inquiry* 38, 671–714.
- Ning, Chunyan. 1993. The overt syntax of relativization and topicalization. *PhD thesis*, University of California at Irvine.
- Nurse, Derek. 2007. 'Did Proto-Bantu have a syntactic or an analytic structure?', in: *School of Oriental and African Studies Working Papers in Linguistics* 15, 239-256.
- Paul, Waltraud. 2008. 'The serial verb construction in Chinese: A tenacious myth and a Gordian knot', in: *Linguistic Review* 25, 2008, 367-411.
- Rizzi, Luigi. 1986. Null objects in Italian and the theory of pro. *Linguistic Inquiry* 17. 501-557.
- Sagart, Laurent. 1999. *The Roots of Old Chinese*. Amsterdam and Philadelphia: Benjamins.
- Sapir, Edward. 1921. *Language. An introduction to the study of speech*. New York: Harcourt, Brace and Company.
- Sinnemäki, Kaius. 2011. Language universals and linguistic complexity. Helsinki: University of Helsinki, Department of Modern Languages.
- Verhaar, John W. M. 1995. *Toward a Reference Grammar of Tok Pisin. An Experiment in Corpus Linguistics*. Honolulu: University of Hawai'i press.
- Wang, Jingling. 2003. *Die Relativkonstruktionen im Chinesischen. Eine diachrone Studie*. München: iudicium [SinoLinguistica, Band 10]. PhD Dissertation, University of Mainz.
- Wang, W. S.-Y. and Cheng, C.-C. 1970. Implementation of phonological change: The Shuang-feng Chinese case. *Chicago Linguistics Society* 6, 552-559.