

Walter Bisang*

Hidden Complexity – The Neglected Side of Complexity and Its Implications

Abstract: Linguistic complexity is the result of the two motivations of explicitness and economy. Most approaches focus on the explicitness-side of complexity (overt complexity) but there is also an economy-oriented side to it (hidden complexity). The aim of the paper is to introduce hidden complexity and to show how it opens new perspectives on central issues of linguistics. It will discuss the following questions: (i) the implementability of hidden complexity for psycholinguistic experiments, (ii) the evolution of linguistic complexity, (iii) its realisation in creole languages and (iv) its consequences for theories on the architecture of grammar. With this aim, the paper is a programmatic paper with the intention of pointing out ideas for future research.

Keywords: architecture of grammar, complexity, creoles, economy, explicitness, maturation, numeral classifiers, pragmatics, radical pro-drop

DOI 10.1515/lingvan-2014-1014

1 Introduction – complexity beyond form

Linguistic complexity is discussed from various perspectives. Typology focuses on the properties of the form by which grammatical categories are expressed and connected (Dahl 2004; Miestamo 2008; Sinnemäki 2011, and others). From the perspective of production and comprehension, complexity is defined in terms of costs and difficulty of acquisition. This second view is adopted in Hawkins' (2004) efficiency-based approach and it is crucial for discussing effects of complexity in language contact and their dependence on the number of bilingual L1 and L2 learners involved (Kusters 2003; Trudgill 2011). A third type of approaches is based on algorithmic information theory. This type is known under the term of Kolmogorov complexity – a measure of the length of description that is needed for specifying an object (Li and Vitányi 2008). In this context, complexity is often measured by compressing a certain object or description by the use of a computer algorithm (e.g. a zip-program as in Juola 2008).¹

Each of these perspectives is based on linguistic form. This is straightforward in the typological view. Typical criteria of form are (i) the number of marked members in a phonemic inventory, (ii) the number of rules that apply within a certain domain of grammar and (c) the fine-grainedness of the semantic/pragmatic distinctions expressed by the linguistic forms belonging to that domain (cf. McWhorter 2001, 2005). Another good indicator is inflection with its idiosyncrasies. In the other two types of complexity, the formal properties of linguistic units are either studied as triggers of effects of costs and difficulties or as elements that need to be integrated into a coherent and economic description.

This paper will argue that the above form-based approaches represent only one side of complexity, the side of overt complexity, that has to do with explicitness. There is another side that is determined by economy and depends on pragmatic inference (Bisang 2009, 2013, 2014a). This other side, which will be

¹ The first type of complexity belongs to the ontological mode of complexity in Sinnemäki's (2011) discussion. The second and the third types belong to the functional mode and the epistemic mode, respectively.

*Corresponding author: Walter Bisang, Department of English and Linguistics (FB 05), University of Mainz, Jakob-Welder-Weg 18, D-55099 Mainz, Germany, E-mail: wbisang@uni-mainz.de

called “hidden complexity”, is independent of overt complexity (cf. Section 4) and opens new horizons in linguistic research. Four of them will be programmatically described in this paper:

1. Since hidden complexity as defined in this paper can be analysed from an incremental perspective, it can be implemented for psycholinguistic experiments and thus offers explanatory potential for typological findings (Section 5).
2. In an evolutionary scenario, there are two types of maturation processes, one of them leads to overt complexity (Dahl 2004), the other one enforces hidden complexity (Section 6).
3. The controversial question of the simplicity of creole grammars in comparison to grammars of older languages as asked by McWhorter (2001, 2005) is based on criteria for overt complexity. A look at these languages from the perspective of overt vs. hidden complexity shows that (i) their grammars do not necessarily allow the production of the most simple overt structures cross-linguistically and that (ii) these complexity properties are related to the complexity properties of the input languages. This has its impact on the extent to which creoles may provide evidence on the evolution of language (Section 7).
4. Given the importance of pragmatic inference as the hidden side of complexity, its integration into assumptions on the architecture of grammar is a challenge to future research. The integration of economy-based complexity may also reconcile Evans and Levinson’s (2009) focus on cross-linguistic diversity with the possibility of theoretical generalizations (Section 8).

Each of these issues will be discussed in a separate section after the definition of hidden complexity (Section 2), two examples (Section 3) and some evidence of the independence of hidden and overt complexity (Section 4).

2 Hidden vs overt complexity

The distinction between hidden and overt complexity is based on the “articulatory bottleneck” as described by Levinson (2000: 6, 27–30). Human speech encoding is by far the slowest part of speech production and comprehension. Other processes like prearticulation, parsing and comprehension show a much higher transmission rate. This creates an asymmetry between articulation and inference in terms of costs – “[i]nference is cheap, articulation expensive” (Levinson 2000: 29). For that reason, every utterance is the result of the two competing motivations of economy vs. explicitness.

The impact of economy and explicitness on the grammar of individual languages is well-known since quite some time (von der Gabelentz 1891: 251; economy vs. iconicity in Haiman 2011; faithfulness vs. markedness constraints in Optimality Theory, e.g. Kager 1999). In the case of complexity, explicitness stimulates overt complexity. It supports the production of grammatical markers which are characterized by their obligatoriness and by fine-grained semantic distinctions. Economy is responsible for “hidden complexity”. It favours grammars (a) in which the speaker is not forced to overtly use grammatical markers and (b) it has multifunctional markers whose concrete meaning must be pragmatically inferred (Bisang 2009).

This paper will only address obligatoriness, which is defined in the sense that the speaker has to express a certain value of a grammatical category in a given syntactic environment (e.g. the expression of past vs. non-past in a finite declarative clause of a language with a binary tense system). Hidden complexity applies if there is an expression format for a given grammatical category with its language-specific system of distinctions (e.g. for tense, person, number, etc.) and if the information concerning that category can be left unmarked if it is possible to infer it pragmatically. Since pragmatic inference follows its own rules, the inferred information belongs to the semantic domain of the category as a whole but it does not necessarily correspond to a specific value reflected by the overt marking system.² This definition does not include

² An example: If two coordinated singular noun phrases in the subject position are represented by a zero pronoun in subsequent clauses, the inference of duality is possible even if the pronominal system of a language does not have a dual.

grammatical categories that are attested somewhere in the languages of the world but happen not to be expressed in a given language. For that reason, there is no hidden complexity with regard to evidentiality in English because evidentiality is not a grammatical category in that language.

There are various ways of measuring hidden complexity. One of them is by assigning numbers to degrees of hidden complexity in individual grammatical domains of individual languages, another one is by using information theoretic measures.³ The assignment of values can be exemplified by adapting the typological method of measuring overt complexity as developed by Parkvall (2008) to hidden complexity. If marking a certain category is obligatory its hidden-complexity could be assigned 0, if it is not obligatory its value could be 1. If necessary, one may assign scores between 0 and 1 (e.g. non-pro-drop = 0, Spanish-type pro-drop = 0.5, radical pro-drop = 1). Since a look at the *World Atlas of Language Structures* (Haspelmath et al. 2005) reveals that optionality is addressed only for three features (nominal plurality, classifiers, pronominal subject), a lot of work is needed for hidden complexity. While methods of assigning complexity values within individual grammatical domains depend on a close look at grammatical structures, information theoretic approaches tend to take a “man-from-mars” perspective on large text corpora.

Two approaches will be discussed for pointing out some problems and potential predictions. Moscoso del Prado (2011) compares the frequency of structure repetition in sequences of characters at different levels between short-range and long-range correlations. As he shows, there is a general tendency across his 16 corpora from typologically different languages to favour repetitions at long scales and to disfavour repetitions at shortest scales. This constant pattern is characteristic of human language. It can be seen as a general reflection of reusing a term once it has been introduced across sentences, while there is a tendency to avoid repetition of structures within a word.⁴ Cross-linguistic variation can be observed in the mid-frequency region that corresponds to intra-sentential structures. In Chinese and English, repetition happens below chance level in that region, while it takes place above chance level in Greek and is neutral in Indonesian, German and Latin (Moscoso del Prado 2011: 3–4). The comparatively low repetition rate in Chinese and English may be due to the limited morphology available in these languages or, in the case of Chinese, (also) to its hidden complexity, which allows the omission of grammatical information that has been introduced earlier (cf. Section 3).⁵ How these two factors interact in Chinese is difficult to say from the data discussed by Moscoso del Prado (2011).

A possible solution to this problem would be the use of parallel texts in English and Chinese, i.e., texts that represent the same content in both languages. In such a pair, one should ideally expect that the intra-sentential frequency of repetition is lower in Chinese than in English, since there are still some obligatory grammatical elements in English (subject/object pronouns, the plural marker -s, finite verb forms for past and agreement -s in present) for which there are no corresponding obligatory markers in Chinese (subject/object pronouns, the plural suffix -men for human nouns, perfective suffix -le). As is pointed out by Juola (2008: 96), the problem with parallel texts is that they are translations from a source text and that translations generally tend to provide more explicit information than the source texts even beyond grammatical categories. This problem could be resolved by using a sample of parallel corpora based on a

³ Since this paper is about more general questions of complexity, it is important to show that hidden complexity can be measured but there will be no concrete complexity calculations.

⁴ An exception is reduplication, which is relatively frequent in Indonesian, one of the corpus languages used by Moscoso del Prado (2011).

⁵ The argument for this effect to be due to hidden complexity runs as follows. If grammatical information that has been introduced earlier does not need to be repeated in subsequent sentences one can assume that there is a considerable number of sentences in which this information is not marked and thus not repeated. Since hidden complexity is only an option of economy provided by the grammar at the level of the sentence, higher principles of text structure may still favour the repetition of grammatical information for reasons of explicitness. For that reason, Chinese tunes into the cross-linguistic preference of repetition at the extra-sentential level like all other languages analyzed by Moscoso del Prado (2011).

source language with high hidden complexity (e.g. Chinese) and another one based on a source language with a relatively high degree of overt complexity. If the target languages are the same and if their hidden complexity is lower than in Chinese, the differences in size between the text in the source language and the translated texts should be higher in the sample based on high hidden complexity than in the overt-complexity sample, since many of the non-obligatory categories in the hidden-complexity sample would have to be explicitly marked in the translations into the languages which require more grammatical explicitness.

3 Two examples

Hidden complexity generates grammars that produce morphosyntactic structures which are very simple from the perspective of overt complexity but need a lot of inferential work for being decoded. This will be illustrated by the examples of radical pro-drop and numeral classifiers in East and mainland Southeast Asian (EMSEA) languages,⁶ which are characterized by their particularly high degree of hidden complexity (Bisang 2008, 2013).

Radical pro-drop is defined by the omission of verbal arguments without the existence of concomitant agreement morphology on the verb (Neeleman and Szendrői 2007). This phenomenon is extremely common in EMSEA languages (Bisang 2014b, 2015). In (1) from Chinese, the subject and the object noun phrase are omitted in (1b) because they can be inferred from (1a):

- (1) a. *Nǐ zuótiān mǎi-le píngguǒ ma?*
 you yesterday buy-PFV apple Q
 ‘Did you buy apples yesterday?’
 b. *Ø mǎi-le Ø.*
 buy-PFV
 ‘[I] bought [some].’

In some EMSEA languages, numeral classifiers express definiteness or indefiniteness if they occur in the bare classifier construction without a numeral [CL N] (Cheng and Sybesma 1999; Li and Bisang 2012). In Wu Chinese (dialect of Fuyang), the classifier marks definiteness with preverbal [CL N] and indefiniteness if [CL N] occurs postverbally (2). In other languages, [CL N] marks both functions irrespective of word order⁷ (Vietnamese, some Sinitic languages) and in a third type it only marks definiteness (cf. (3) on Hmong).

- (2) Wu Chinese (Li and Bisang 2012: 336):
 $k\gamma$ *lɔpan ma lə bu ts^hots^hɿ.*
 CL boss buy PFV CL car
 ‘[The boss/*a boss] bought [a car/*the car].’

Classifiers as markers of (in)definiteness are not obligatory and thus differ significantly from article systems (Lyons 1999). They are optional with unique nouns (Li and Bisang 2012) and they are omitted in contexts of

⁶ The families involved are Sinitic (Sino-Tibetan), Mon-Khmer (Austroasiatic), Tai (Tai-Kadai), Hmong-Mien and Chamic (Austronesian).

⁷ In languages of this type, the interpretation of [CL N] in terms of (in)definiteness depends on factors like event type, aspect marking and noun type (relational vs. non-relational noun, unique vs. non-unique noun) as well as on information structure and context. The details of [CL N] interpretation in languages of this type need more research.

anaphoric definite reading. In (3) from Hmong, *plis* “leopard” appears as a bare noun in line 4 after its referential status is obvious from previous context:

(3) Hmong (Mottin 1980):

<i>Muaj</i>	<i>ib</i>	<i>tug</i>	<i>plis</i>	<i>caum</i>	<i>los.</i>	<i>Mas</i>	<i>nws</i>	<i>txawm</i>
there.is	one	CL	leopard	chase	come	then	he	then
<i>tib hnee</i>		<i>rau</i>	<i>tus</i>	<i>qaib</i>	<i>qus</i>	<i>nti</i>		<i>ceev ceev.</i>
shoot.with.crossbow		at	CL	chicken	wild	toss.to.and.fro		quickly
<i>Mas</i>	<i>tus</i>	<i>plis</i>	<i>txawm</i>	<i>dhia</i>	<i>dhia</i>	<i>los</i>	<i>khawm nkaus</i>	<i>thiab</i>
Then	CL	leopard	then	jump	RED	come	catch	and
<i>zaum tsawg</i>	<i>noj.</i>	<i>Plis</i>	<i>muab</i>	<i>noj</i>	<i>noj</i>	<i>tag.</i>		
sit.down		eat	leopard	take	eat	RED	complete	

‘There was **a leopard** chasing after them [i.e. the wild chicken that have been introduced earlier]. Then he [= Qua Tu, the protagonist] shot at the group of wild chicken and they were tossing to and fro. Then, **the leopard** jumped towards [them], caught [them], sat down and ate [them]. **The leopard** ate up all of them.’

4 The independence of hidden and overt complexity

Economy is about minimizing articulation, while explicitness aims to specify the intended meaning of an utterance for reducing the number of possible interpretations. Due to these differences in function, trade-offs between hidden and overt complexity are unlikely. In spite of this, the degree to which hidden and overt complexity can develop in grammars is limited. This is motivated by the properties of the parser in online processing. Even though explicitness or overt complexity supports processing to a certain extent by providing cues that enhance the incremental analysis of the parser, too much overt information makes the analysis more demanding and thus creates costs. Economy minimizes the risk of creating disturbances through too much information but if too much information is missing this produces costs as well, since the parser will need time to decide between too many alternative analyses.

In concrete utterances, both principles operate and restrict each other during online speech analysis and speech production at any moment. In principle, each position can be seen as an option for expressing or omitting grammatical information according to the morphosyntax of a given language. In some positions the grammar is open to choices between economy and explicitness, in others a given category has to be expressed or needs to be omitted (also cf. Section 5 on incrementality). In the course of time, restrictions of obligatoriness may change. Often, these changes are triggered by morphophonological processes of erosion, in which a given marker may eventually disappear completely. The constant competition between the two principles in their interaction with the parser limit the processes of linguistic change and create a certain medium area (cf. Figure 1⁸) between overt and hidden complexity (also cf. Section 6 on maturation). Cross-linguistic variation is situated somewhere within that medium area – it can be seen as the result of successful solutions to this competition.

Another look at radical pro-drop and (in)definiteness shows that these two grammatical domains combine hidden complexity with high overt complexity in some EMSEA languages. This proves that the two sides of complexity are independent.

⁸ The arrows in Figure 1 indicate increasing values for overt and hidden complexity as well as for costs.

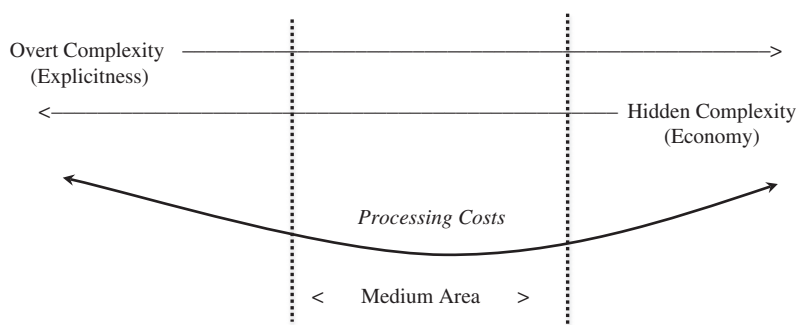


Figure 1 The medium area between hidden and overt complexity

The pronominal systems of some EMSEA languages are quite complex even if they are not obligatory. The system of Hmong consists of nine non-agglutinative forms combining number (singular, dual, plural) and person (1st/2nd/3rd):

(4) The system of Hmong pronouns:

	SG	DUAL	PL
1. person	<i>kuv</i>	<i>wb</i>	<i>peb</i>
2. person	<i>koj</i>	<i>neb</i>	<i>nej</i>
3. person	<i>nws</i>	<i>nkawd</i>	<i>lawv/luag</i>

The overt complexity of (in)definiteness marking by [CL N] is materialised in the rich inventory of semantic distinctions associated with classifiers. Nouns in EMSEA languages typically select classifiers according to the semantic properties of the head noun. Important criteria are features like [human], [animate], [shape], [consistency], [size], [location] and [man-made object].⁹ These fine-grained distinctions clearly contribute to overt complexity. Even the classifier system of Weining Ahmao, which probably scores highest for overt complexity among EMSEA languages, still participates in hidden complexity because it is not fully obligatory. In the singular, Weining Ahmao classifiers inflect for definite/indefinite and size (augmentative, medium, diminutive) and thus occur in six different morphological forms (Gerner and Bisang 2008).

The independence of hidden vs. overt complexity is less obvious in well-known Indo-European languages like English or Spanish. They have to use overt marking for the features of person, number and gender of the subject (English is non-pro-drop, while Spanish is pro-drop but with obligatory agreement) and they have articles which cannot be omitted even if referential status is established in discourse. It would be interesting to see how the ratio of hidden vs. overt complexity is in languages with more complex overt systems.¹⁰ Given the lack of typologically reliable data on hidden complexity (cf. Section 2), this needs a lot of data collection.

5 Incrementality

Hidden complexity cannot only be measured (Section 2), it can also be compared to effects of overt complexity in contexts of incrementality. The use of \emptyset vs. pronoun in radical pro-drop languages and the use of [CL N] vs. [N] are expected to trigger different pragmatic effects. In the case of radical pro-drop as described for Chinese by Huang (1994), the use of \emptyset implies coreference with a default argument

⁹ There is a rich literature on classification. I only mention the classic study of Allan (1977).

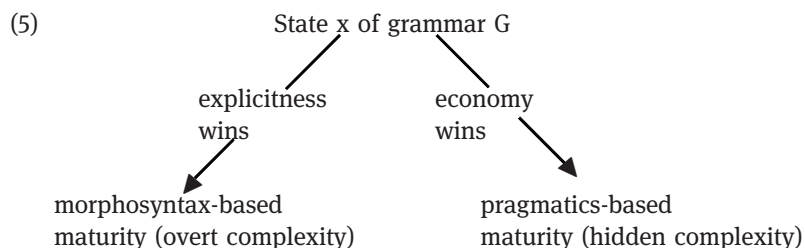
¹⁰ E.g. in the languages that score highest in Parkvall (2008: 274–5): Burushaski, Copainalá Zoque, Khoekhoe (Spanish is on rank 11, English on rank 78 of 155 languages).

(I-implicature), while the use of the more marked pronoun *tā* “s/he” favours the complementary interpretation of coreference with a non-default argument (M-implicature).¹¹ Since the predictions on potential NP arguments vary at the position of { \emptyset vs. pronoun}, this position can be implemented for experiments of online processing in which hidden vs. overt complexity is predicted to show different effects in specific contexts.

Miestamo (2008) points out the difference between relative complexity that depends on the perspective from which a linguistic structure is looked at (e.g. speaker, hearer, L1 or L2 learner) and absolute complexity that does not depend on a particular perspective. In his view, only absolute complexity is suited for typological generalizations because it is independent of any particular perspective. Even though this may look like a clear-cut dichotomy between typological approaches and cost-based approaches, the two aspects of complexity are related. In fact, the structures looked at from a typological perspective may be explained by cost-related factors. The relation between the two aspects can be illustrated by another look at incrementality. When the human parser arrives at a given marker with the feature X in an utterance Y that feature triggers certain effects in the human brain. The same feature will also be observed by a linguist who analyses utterance Y from a holistic and absolute perspective. From such a perspective, it is to be expected that typological findings from the absolute perspective are motivated by different predictions at a given position in online processing.

6 The evolution of hidden vs overt complexity

The complexity of individual languages changes through time. In Dahl’s (2004: 103–5) model of maturation, complexity increases through time. It is the result of the accumulation of grammatical material in the grammar G of a language that was not present at an earlier stage G’ of that language (also cf. Ansaldo and Nordhoff 2009: 358). Thus, phenomena like complex word structure (e.g. inflectional morphology), lexical idiosyncrasy (e.g. inflectional classes) and syntactic phenomena that are dependent on inflectional morphology (e.g. agreement)¹² are products of an evolutionary process of significant length. With these phenomena, Dahl’s (2004) concept of maturation focuses on the emergence of properties associated with overt complexity as it is motivated by explicitness. From the perspective of economy and pragmatic inference, there is another 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.



Seen from the perspective of constant competition between economy and explicitness at any position in an utterance (cf. Section 4), the obligatorification of a category in a given morphosyntactic position may

¹¹ The two implicatures are defined in terms of Levinson’s (2000) neo-Gricean approach. The I-implicature (or Informativeness Principle) is based on the idea that “[w]hat is expressed simply is stereotypically exemplified” and is related to Grice’s (1975) second Maxim of Quantity “[d]o not make your contribution more informative than required”. The M-implicature (or Manner Principle) is related to Grice’s Maxim of Manner (e.g. “avoid prolixity”). Thus, the use of a more elaborate expression suggests that the stereotypical interpretation does not apply.

¹² Cf. Dahl (2004: 114–5) for the whole list.

enhance additional overt complexity elsewhere at a next stage if that category is exploited for additional functions (e.g. noun class markers as markers of agreement within the noun phrase) or if its presence necessitates additional rules (e.g. word order, simultaneous presence of another category). This type of increasing overt complexity through time is paralleled in economy-based maturation. In the case of EMSEA languages, the members of the five language families involved were characterized by reduced morphology that typically did not express inflectional categories as early as about 2,000 years ago (Bisang 2014b). Later on, the interaction of a set of factors preserved this situation and further enhanced hidden complexity over a long period of time. The factors involved are described in Bisang (2008, 2014b) and include language contact, phonological properties that withstand the erosion of grammatical markers, a considerable degree of non-obligatoriness that is further developed and the multifunctionality of many grammatical markers. As is to be expected with competing motivations, even if economy is remarkably dominant in EMSEA languages there are also some instances in which explicitness wins. This is the case with (i) the obligatoriness of classifiers with numerals, (ii) two different verb forms for expressing tense-aspect distinctions in various Sinitic languages (Arcodia 2013) and the inflection of numeral classifiers in Weining Ahmao (Section 4).

7 Creole languages

In a scenario of competition between overt and hidden complexity, the crucial question is to what extent the grammar of a language does allow the omission of overt information if it can be pragmatically inferred. This section will argue that many creole grammars force their speakers to use overt marking where grammars of older languages spoken in EMSEA are open to hidden complexity. Thus, the grammars of creole languages do not necessarily allow the production of the cross-linguistically most simple overt structures. While the high degree of hidden complexity is the result of economy-based maturation in EMSEA languages, the lack of hidden complexity in the creoles concerned seems to be due to the linguistic input in the early phase of pidginization.

The case of pro-drop can be addressed with data from *The Atlas of Pidgin and Creole Language Structures Online (APiCS)* (Haspelmath et al. 2013) on feature 62 “Expression of pronominal subjects”. In a sample of 76 languages, 18 (23.7%) belong to the category “[o]ptional pronoun words”, which corresponds to radical pro-drop given the absence of person agreement on the verb. This is a relatively small percentage that shows that many pidgins and creoles cannot produce simple overt structures like (1b) from Chinese.

The marking of definiteness is characterized by extensive variation. Déprez (2007) compares the functional range of bare nouns in four creoles and concludes that “there is no prototypical creole determiner system” (Déprez 2007: 319). While bare nouns are possible in contexts of anaphoric definite reading in Seychelles Creole, they are not accepted in Mauritian and Haitian Creole. In Cape Verdean Creole, the functions of bare nouns depend on animacy and on the presence of overt plural marking. In a somewhat more extensive list of 15 creoles, Baptista (2007: 467) presents 7 languages with bare nouns that can express specificity and non-specificity (among them Cape Verdean Creole, for which this is only true with non-animate and with plural-marked nouns).

The above two examples indicate that there is considerable variation in the ability to create simple surface structures in creoles. In more recent studies, the relevance of complexity is sometimes limited to subtypes of complexity or it is denied completely. Bakker et al. (2011) straightforwardly argue that complexity is not the right property for distinguishing creoles from other languages. A closer look at contact-based models shows that this statement may need some specification concerning different types of complexity. Thus, Good (2012) claims that creoles show high degrees of simplicity only in complexity that derives from paradigmatic complexity, i.e., the range of subdistinctions available, while their syntagmatic complexity as it is reflected in the structure of a given linguistic object is average. This is due to the fact that paradigmatic complexity implies the transfer of a whole set of markers at the initial stage of pidgin formation

(jargonization in terms of Good 2012), while syntagmatic complexity is limited to the transmission of only one marker and thus needs much less coordinative effort. With the specification of paradigmatic vs. syntagmatic complexity, only low paradigmatic complexity can serve as a potential criterion for identifying creoles.

The distinction between hidden vs. overt complexity leads to the modification of another important assumption concerning the impact of contact on creole structures. In McWhorter's view (2001: 126), the absence of many properties contributing to the overt complexity of older languages in creoles is due to the initial situation of extreme contact in which the creators of creoles "strongly tended to eschew traits from their native languages which were incidental to basic communication". The fact that many creoles are not radical pro-drop (cf. above) suggests that there are limitations to the extent to which creole creators "eschew traits from their native languages". As was argued in Bisang (2013), the probability of omitting information concerning certain grammatical categories in creole languages considerably depends on whether these categories are obligatory in the grammars of the input languages. If the input languages allow radical pro-drop, its speakers will also be ready to drop that information in situations of extreme contact from which later creoles develop. If the input languages are not radical pro-drop as in the case of creoles with West African substrates and European superstrates, their speakers seem to feel the need of overtly providing information concerning the subject even in situations of extreme contact that is characteristic of creoles. Thus, creoles with such input languages tend to lack radical pro-drop. As this discussion shows, extreme contact situations may undo the results of earlier processes of explicitness-based and economy-based maturation in the input languages involved, but with certain limitations. If creole structures are basically the result of language contact it remains questionable to what extent creole languages may provide direct insights into the evolution of language. At least, McWhorter's (2001: 126) statement that "creoles are unique in reflecting the innate component of the human language capacity more closely than older languages do" needs refinement.

8 Theoretical consequences

Complexity cannot only be discussed in terms of properties of languages but also in terms of the design of linguistic theories. As Culicover and Jackendoff (2005, 2006) and Culicover (2013) point out, hypotheses on the syntax-meaning relation have their impact on assumptions concerning the complexity of the architecture of grammar. If one adopts the syntactocentric view of the Chomskyan model, semantics cannot have a more elaborate structure than the syntax that expresses it. This is due to the basic assumption that "[t]he meaning of a compound expression is a function of the meaning of its parts and of the syntactic rules by which they are combined" (Partee, terMeulen and Wall 1990: 30).

In non-syntactocentric theories, among them *Simpler Syntax Hypothesis (SSH)* (Culicover and Jackendoff 2005, 2006), syntax is not the only generative component of grammar, it is a mediator between the other two components of phonology and semantics and it is "only as complex as it needs to be to establish interpretation" (Culicover and Jackendoff 2006: 413). Thus, "[t]he most explanatory syntactic theory is one that imputes the minimum structure necessary to mediate between phonology and meaning" (Culicover and Jackendoff 2005: 5).

If the complexity of syntax is reduced, the burden of establishing adequate interpretation is partly transferred to the semantic component. A short look at radical pro-drop will reveal whether this works for hidden complexity. In Culicover and Jackendoff's (2005: 195–7) approach, there is a special system in language that concerns the realization of the NP arguments of a verb, which is called the "Grammatical Function tier" (GF-tier). In the case of non-pro-drop, this tier is linked to both semantics and syntax. In the case of pro-drop with concomitant verb agreement, the GF-tier is only linked to semantics or Conceptual Structure. Problems arise with radical pro-drop. Since the subject cannot be retrieved from semantics, it has to be inferred pragmatically. Thus, a pragmatic component needs to be integrated into the architecture of

grammar to account for radical pro-drop (and other grammatical categories whose marking is not obligatory in a given language, as for instance (in)definiteness or number in EMSEA languages). This does not only apply to models consisting of more than one component like SSH, it also applies to syntactocentric approaches if there are problems of licensing omitted grammatical categories as in the case of radical pro-drop. As was pointed out by Neeleman and Szendrői (2007), there are theoretical problems with the topic-drop approach (Huang 1984) as well as with the blocking-by-determiners approach (Speas 1994). Neeleman and Szendrői's (2007) own explanation that radical pro-drop requires agglutinating morphology on pronouns does not work with a radical pro-drop language like Hmong (cf. the pronominal system in (4)).¹³

The above discussion shows that economy requires the integration of pragmatics into models of the architecture of language. Of particular interest in this context are questions concerning the grammatical categories involved. What grammatical categories are concerned with the competition between economy and explicitness? To what extent do these categories form a set of cross-linguistically recurrent core categories? With such an approach, it should be possible to focus on cross-linguistic structural differences as postulated by Evans and Levinson (2009) but it should also be possible to see which categories are of central theoretical importance and which ones are the result of extensive maturation based on conditions provided by earlier stages of maturation.

9 Conclusion

The present paper introduced hidden complexity and its independence of overt complexity. It showed how this new perspective can be used for processing experiments and how it affects central issues concerning the evolution of complexity, the discussion of creole languages and theories about the architecture of grammar.

References

- Allan, K. 1977. *Classifiers*. *Language* 53. 285–311.
- Ansaldo, U. & S. Nordhoff. 2009. Complexity and the age of languages. In E. O. Aboh & N. Smith (eds.), *Complex processes in new languages*, 345–363. Amsterdam/Philadelphia, PA: Benjamins.
- Arcodia, G. F. 2013. Grammaticalization with coevolution of form and meaning in East Asia? Evidence from Sinitic. *Language Sciences* 40. 148–167.
- Bakker, P., A. Daval-Markussen, M. Parkvall & I. Plag. 2011. Creoles are typologically distinct from non-creoles. *Journal of Pidgin and Creole Languages* 26. 5–42.
- Baptista, M. 2007. Properties of noun phrases in creole languages. A synthetic comparative exposition. In M. Baptista & J. Guéron (eds.), *Noun phrases in Creole languages. A multi-faceted approach*, 461–470. Amsterdam/Philadelphia, PA: Benjamins.
- Bisang, W. 2008. Grammaticalization and the areal factor: The perspective of East and mainland Southeast Asian languages. In M. J. López-Couso & E. Seoane (eds.), *Rethinking grammaticalization. New perspectives*, 15–35. Amsterdam/Philadelphia, PA: Benjamins.
- Bisang, W. 2009. On the evolution of complexity—Sometimes less is more in East and mainland Southeast Asia. In G. Sampson, D. Gil & P. Trudgill (eds.), *Language complexity as an evolving variable*, 34–49. Oxford: Oxford University Press.
- Bisang, W. 2013. Language contact between geographic and mental space. In P. Auer, M. Hilpert, A. Stukenbrock & B. Szendrői (eds.), *Linguistic perspectives on space: Geography, interaction, and cognition*, 61–100. Berlin: Mouton de Gruyter.
- Bisang, W. 2014a. Overt and hidden complexity—Two types of complexity and their implications. *Poznan Studies in Contemporary Linguistics* 50(2). 127–143.

¹³ For some more explanations of problems with existing accounts of radical pro-drop, cf. Bisang (forth, Section 3) on Neeleman and Szendrői (2007).

- Bisang, W. 2014b. On the strength of morphological paradigms—A historical account of radical pro-drop. In M. Robbeets & W. Bisang (eds.), *Paradigm change in historical reconstruction: The transeurasian languages and beyond*, 23–60. Amsterdam/Philadelphia, PA: Benjamins.
- Bisang, W. 2015. Problems with primary vs. secondary grammaticalization: The case of East and mainland Southeast Asian languages. *Language Sciences* 47. 132–147.
- Cheng, L.-L. S. & R. Sybesma. 1999. Bare and not so bare nouns and the structure of NP. *Linguistic Inquiry* 30. 509–542.
- Culicover, P. W. 2013. *Grammar and complexity. Language at the intersection of competence and performance*. Oxford: Oxford University Press.
- Culicover, P. W. & R. Jackendoff. 2005. *Simpler syntax*. Oxford: Oxford University Press.
- Culicover, P. W. & R. Jackendoff. 2006. The simpler syntax hypothesis. *Trends in Cognitive Sciences* 10(9). 413–418.
- Dahl, Ö. 2004. *The growth and maintenance of linguistic complexity*. Amsterdam/Philadelphia, PA: Benjamins.
- Déprez, V. 2007. Implicit determination and plural. In M. Baptista & J. Guéron (eds.), *Noun phrases in Creole languages. A multi-faceted approach*, 301–336. Amsterdam/Philadelphia, PA: Benjamins.
- Evans, N. and S. C. Levinson. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32. 429–448.
- Gabelentz, G. von der. 1891. *Die Sprachwissenschaft, ihre Aufgaben, Methoden und bisherigen Ergebnisse*. Leipzig: T.O. Weigel Nachfolger.
- Gerner, M. & W. Bisang. 2008. Inflectional speaker-role classifiers in Weining Ahmao. *Journal of Pragmatics* 40. 719–731.
- Good, J. 2012. Typologizing grammatical complexities or why creoles may be paradigmatically simple but syntagmatically average. *Journal of Creole and Pidgin Languages* 27. 1–47.
- Grice, P. H. 1975. Logic and conversation. In P. Cole & J. L. Morgan (eds.), *Speech acts*, 41–58. New York: Academic Press.
- Haiman, J. 2011. Competing motivations. In J. J. Song (ed.), *The Oxford handbook of linguistic typology*, 148–165. Oxford: Oxford University Press.
- Haspelmath, M., M. S. Dryer, D. Gil & B. Comrie (eds.). 2005. *The world atlas of language structures*. Oxford: Oxford University Press.
- Haspelmath, M. and the APiCS Consortium. 2013. Expression of pronominal subjects. In S. M. Michaelis; P. Maurer, M. Haspelmath & M. Huber (eds.), *Atlas of Pidgin and Creole language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. <http://apics-online.info/parameters/62> (accessed 15 July 2014).
- Hawkins, J. A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
- Huang, C.-T. J. 1984. On the distribution and reference of empty pronouns. *Linguistic Inquiry* 15. 531–574.
- Huang, Y. 1994. *The syntax and pragmatics of anaphora*. Cambridge: Cambridge University Press.
- Juola, P. 2008. Assessing linguistic complexity. In M. Miestamo, K. Sinnemäki & F. Karlsson (eds.), *Language complexity: Typology, contact, change*, 89–108. Amsterdam/Philadelphia, PA: Benjamins.
- Kager, R. 1999. *Optimality theory*. Cambridge: Cambridge University Press.
- Kusters, W. 2003. *Linguistic complexity. The influence of social change on verbal inflection*. PhD Dissertation. University of Leiden.
- Levinson, S. C. 2000. *Presumptive meanings. The theory of generalized conversational implicatures*. Cambridge, MA: The MIT Press.
- Li, M. & P. M. B. Vitányi. 2008. *An introduction to Kolmogorov complexity and its applications*, 3rd edn. New York: Springer.
- Li, X. & W. Bisang. 2012. Classifiers in Sinitic languages: From individuation to definiteness-marking. *Lingua* 122. 335–355.
- Lyons, Christopher. 1999. *Definiteness*. Cambridge: Cambridge University Press.
- McWhorter, J. H. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology* 5. 125–166.
- McWhorter, J. H. 2005. *Defining Creole*. Oxford: Oxford University Press.
- Miestamo, M. 2008. Grammatical complexity in a cross-linguistic perspective. In M. Miestamo, K. Sinnemäki & F. Karlsson (eds.), *Language complexity: Typology, contact, change*, 23–41. Amsterdam/Philadelphia, PA: Benjamins.
- Moscato del Prado, Fermin. 2011. The universal “shape” of human languages: Spectral analysis beyond speech. Available from Nature Precedings. <http://precedings.nature.com/documents/6097/version/2>
- Mottin, J. 1980. *Contes et légendes Hmong Blanc*. Bangkok: Don Bosco Press.
- Neeleman, A. & K. Szendrői. 2007. Radical pro drop and the morphology of pronouns. *Linguistic Inquiry* 38. 671–714.
- Parkvall, M. 2008. The simplicity of creoles in a cross-linguistic perspective. In M. Miestamo, K. Sinnemäki & F. Karlsson (eds.), *Language complexity: Typology, contact, change*, 265–285. Amsterdam/Philadelphia, PA: Benjamins.
- Partee, B. H., A. ter Meulen & R. E. Wall. 1990. *Mathematical methods in linguistics*. Dordrecht: Kluwer.
- Sinnemäki, K. 2011. *Language universals and linguistic complexity*. Helsinki: University of Helsinki, Department of Modern Languages.
- Speas, M. 1994. Null arguments in a theory of economy of projections. *University of Massachusetts Occasional Papers in Linguistics* 17. 179–208.
- Trudgill, P. 2011. *Sociolinguistic typology. Social determinants of linguistic complexity*. Oxford: Oxford University Press.

Note: The research presented in this paper was supported by the Gutenberg Research College of the University of Mainz.