Grammaticalization and contact—hidden complexity in the languages of East and mainland Southeast Asia

0. Basic idea

- Contact among speakers of different languages (linguistic contact) may induce structural convergence. There are certain areas, called linguistic areas (Sprachbund), in which convergence covers a larger number of languages (also cf. Bisang 2006 on zones of structural convergence).
- East and mainland Southeast Asia (EMSEA) is a linguistic area/zone of convergence.
- The area of EMSEA is characterized by specific properties that affect grammaticalization:
  - Lack of coevolution of meaning and form, prominence of pragmatic inference
  - Grammaticalization processes in EMSEA languages illustrate particularly well
    - the independence of hidden and overt complexity
    - the existence of two types of maturation: explicitness-driven and economy-driven maturation

1. Basic notions

1.1. Language contact and linguistic areas/zones of structural convergence¹

Contact among speakers of different languages has its impact on the structures of the languages involved:

=> structural convergence.

There are many factors that determine the properties and the extent of structural convergence. Thomason (2001) discusses the following factors:

- Bilingualism (2L1 speakers)
- Second language acquisition (L2 speakers)
- Code-switching (?)
- Code alternation
- Passive familiarity
- Negotiation
- Deliberate decisions

¹ As I tried to show in Bisang (2006), it is impossible for various reasons to come up with a precise definition of linguistic areas. In spite of this, a lot of structural convergence through language contact in languages of unrelated/remote genetic origin can be observed. For that reason, I suggested the more neutral term of “zone of convergence”.
Each of these factors depends on the following basic conditions (Thomason & Kaufman 1988):

- Intensity of contact: duration, percentage of population involved, direct/indirect contact
- Availability and number of bilinguals, percentage of second-language learners

One possible outcome of contact-induced convergence are linguistic areas/zones of convergence (German Sprachbund, PL. Sprachbünde). There are various definitions of what makes a linguistic area. I only quote two of them (for a more extensive discussion, cf. Stolz 2002, Bisang 2006):

A linguistic area is “an area which includes languages belonging to more than one family but showing traits in common which are found not to belong to the other members of (at least) one of the families” (Emeneau 1956: 16, note 28).

... a linguistic area is a geographical region containing a group of three or more languages that share some structural features as a result of contact rather than as a result of accident or inheritance from a common ancestor. (Thomason 2001: 99)

1.2. East and mainland Southeast Asian languages (EMSEA) as a zone of convergence

The languages of EMSEA are characterized by their long and highly complex contact history with many different situations of contact at different times and they share a lot of structural similarities. For that reason, they are called a linguistic area by many researchers (Bisang 1996, Enfield 2003, 2005).

The language families spoken in East and mainland Southeast Asia:

- Sino-Tibetan
- Mon-Khmer (a subfamily of Austroasiatic)
- Tai (the core group of the Tai-Kadai languages)
- Hmong-Mien (also called Miao-Yao)
- Chamic (Malayo-Polynesian subfamily of Austronesian).

The families forming the core of East and mainland Southeast Asia as a zone of contact-induced convergence (cf. Bisang 1996, Enfield 2005):

- Mon-Khmer
- Tai
- Hmong-Mien
- Sinitic (“dialects” of Chinese)

1.3. A remark on EMSEA languages and the notion of grammaticalization areas

Heine & Kuteva (2005) describe the phenomenon of contact-induced grammaticalization:

(1) Contact-induced grammaticalization (Heine & Kuteva 2005: 81):
   a. Speakers notice that in language M there is a grammatical category Mx.
   b. They create an equivalent category Rx in language R on the basis of the use patterns available in R.
   c. To this end, they draw on universal strategies of grammaticalization, using construction Ry in order to develop Rx.
   d. They grammaticalize Ry to Rx. [M = Model language, R = Replica language]
Processes of contact-induced grammaticalization take place between two languages but they can also take place across a larger number of languages within a certain coherent geographic area. Heine & Kuteva (2005, chapter 5.2) discuss the following grammaticalization areas:

- The Balkans
- Meso-America
- South Asia
- Southeast Asia
- Ethiopia
- Others: East Anatolia, Balto-Finnic contact area, Balto-Slavic contact area, Daly River area (Australia), Pacific Northwest Coast.

In my view, the situation in EMSEA languages is more complex:

- Contact-induced grammaticalization does play a certain role
- but important aspects of the specific properties of grammaticalization within an area of grammaticalization must have been supported by the properties which the member languages brought with them when they got in contact and which were further developed in the course of time (cf. handout 4).
- The two most important properties are (Bisang 2004):
  - High relevance of pragmatic inference
  - Limited of coevolution of meaning and form
    (cf. Lehman 1995 on autonomy and § 2.1 below)

In this lecture, I will focus on pragmatic inference (but I will also briefly discuss the lack of meaning/form coevolution, cf. section 2.1):

The high relevance of pragmatic inference in EMSEA languages, offers the following insights into how complexity can be realised in the world’s languages:

(i) independence of overt vs. hidden complexity => § 1.4
(ii) two types of maturation => § 1.4

Thus, contact and areality can have more important effects than individual instances of structural convergence.

---

2 Area: Large part of Mexico, Guatemala, Belize, El Salvador plus the Southern parts of Honduras, Nicaragua and Costa Rica. Features: (a) attributive possession of the type *his-dog the man*, (b) relational nouns, (c) vigesimal numeral system and (d) non-verb final basic word order, to which absence of switch-reference is correlated (Heine & Kuteva 2005: 199-202).
1.4. On the independence of overt vs. hidden complexity (cf. handout 1, p. 14)

![Diagram showing the independence of overt vs. hidden complexity]

Overt Complexity (Explicitness) \[\text{Processing Costs} \rightarrow \text{Hidden Complexity (Economy)}

\[< \quad \text{Medium Area} \rightarrow \]

Figure 1

1.5. On two types of maturation (cf. handout 1, p. 8)

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.

(2) State \(x\) of grammar \(G\)

\[\text{explicitness wins} \quad \text{economy wins}\]

morphosyntax-based maturity (overt complexity) \quad \text{pragmatics-based maturity (hidden complexity)}

Like explicitness-driven maturation, economy-driven maturation also needs time. 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 2013; also cf. handout 4).

2. Characteristics of grammaticalization in EMSEA languages

2.1. Short remarks on the limited coevolution of form and meaning (cf. handout 2, p. 4 – 7)

The coevolution of form and meaning is quite limited in EMSEA languages (Bisang 2004, 2011). This can be seen from looking at Lehmann’s (1995) criteria for measuring the degree of autonomy of linguistic signs as summarized in the following Table:
Table 1

<table>
<thead>
<tr>
<th></th>
<th>Paradigmatic</th>
<th>Syntagmatic</th>
</tr>
</thead>
<tbody>
<tr>
<td>weight</td>
<td>Integrity (Ia)</td>
<td>Structural scope (Ib)</td>
</tr>
<tr>
<td>cohesion</td>
<td>Paradigmaticity (Iia)</td>
<td>Bondedness (IIb)</td>
</tr>
<tr>
<td>variability</td>
<td>Paradigmatic variability (IIIa)</td>
<td>Syntagmatic variability (IIIb)</td>
</tr>
</tbody>
</table>

**Ia. Integrity (paradigmatic weight)**

Linguistic signs change their semantic weight (they take on a more abstract grammatical function) but their phonological integrity only changes to a limited extent.

There are some exceptions to this rule:

- The Chinese perfective marker -le is derived from the verb lào 了 ‘finish’. Similarly, the continuative marker -zhe is derived from zháo 着 ‘touch’.
- In some Sinitic varieties mostly spoken in the north of China, some initial stages of morphological paradigms can be observed (cf. Arcodia (2013)).

**Ib. Scope (syntagmatic weight)**

The structural size of the construction into which a given grammatical marker is integrated is usually at a phrasal level. Apart from some exceptions like the perfective suffix -le in Mandarin Chinese, grammatical markers do not operate on the word level.

**IIa. Paradigmaticity (paradigmatic cohesion)**

EMSEA languages only show a limited degree of paradigmaticization. They rarely develop morphological paradigms (but cf. (2) above).

**IIb. Bondedness (syntagmatic cohesion)**

Higher degrees of fusion as they are found in cliticization or agglutination are rare in EMSEA languages (cf. again Mandarin –le, Arcodia 2013).

**IIIa. Paradigmatic variability**

High degrees of grammaticalization are generally linked to obligatoriness. In EMSEA languages, only very few grammatical markers are obligatory. A well-known exception are numeral classifiers in the context of counting (cf. § 3.1).

**IIIb. Syntagmatic variability**

Rigid word-order rules are by far the most reliable indicators of high degrees of grammaticalization.
2.2. High relevance of pragmatic inference

The high importance of pragmatic inference is a general property of EMSEA languages that is not limited to grammaticalization.

Huang Y. (1994) claims that the strength of syntax vs. pragmatics is subject to cross-linguistic variation:

There seems to exist a class of language (such as Chinese, Japanese and Korean) where pragmatics appears to play a central role which in familiar European languages (such as English, French and German) is alleged to be played by grammar. In these ‘pragmatic’ languages many of the constraints on the alleged grammatical processes are in fact primarily due to principles of language use rather than rules of grammatical structure.

(Huang 1994: xiv)

LaPolla (2003) looks at cross-linguistic variation in a similar way but he starts out from a different perspective on the relation between context and disambiguation:

Contrary to most work in pragmatics which assumes that context disambiguates language, I am proposing that it is not context that disambiguates language, but language that disambiguates the context of interpretation. That is, rather than assume that the form of the utterance is given and the context is the manipulable variable, we should recognise that in real-world communicative situations, speakers have no choice in terms of the context they are communicating in, but as communicators have choices in terms of the form of the utterance they use. (LaPolla 2003: 114)

The freedom of choice, however, differs across languages:

The point is that languages differ quite a lot in how much they constrain the search for the most relevant interpretation, and in what aspects they chose to constrain. (LaPolla 2003: 134)

No matter which analysis is more adequate, pragmatic inference is more prominent in Chinese and EMSEA languages than in many other languages (among them the languages of Europe). We shall briefly look at two examples:

1. Lack of subject/object-asymmetry in the coordinate-clause construction:


a. [That man]i dropped the watermelon on the ground and ∅i got flustered.
b. ??? [That man]i dropped the watermelon on the ground and ∅i burst.

(4) Chinese: Coordinate-clause construction (LaPolla 1993):

a. 那个人把西瓜掉在地上, 慌了。
   na ge ren ba xigua diao zai di-shang ∅i huang le
   that CL man take watermelon drop to ground-on flustered PF
   ‘[That man], dropped the watermelon on the ground and ∅i got flustered.’

b. 那个人把西瓜掉在地上, 碎了。
   na ge ren ba xigua i diao zai di-shang ∅i sui le
   that CL man take watermelon drop to ground-on break PF
   ‘That man dropped [the watermelon], on the ground and [it], burst.’
2. Governing Categories and the influence of pragmatics

Chinese disrespects the classical Binding Theory in which the reflexive must be bound locally (i.e., within the Governing Category, cf. (9)). In the following example, the subordinate clause has the function of the Governing Category. If the classical Binding Theory is correct, the reflexive pronoun can only refer to the subject of the subordinate clause:

(5) John, believes that [Peter, loves himself\_ij].

Compare Chinese:

(6) Chinese (Huang 1994: 97):

王先生以為須小姐愛上了自己。

\[Wang \quad Xiansheng\_j \quad yiwei \quad [Xu \quad Xiaojie\_i \quad aishang-le \quad ziji\_ij].\]

Wang Mister think Xu Miss fall.in.love-PFV SELF

a. Mister Wang thinks that Miss Xu has fallen in love with herself.

b. Mister Wang thinks that Miss Xu has fallen in love with him.

Thus, the coreference of the reflexive depends on pragmatic inference rather than on Binding Theory, which is a syntactic rule. As the following example shows, binding may depend on the semantics of the verbs involved and the pragmatic inferences that follow from them. In (7), binding is determined by the meaning of the verb 娶 qù ‘marry a woman’ (‘marry a man is 嫁 jià):

(7) Chinese (Huang 1994: 183):

須小姐希望王先生娶自己。

\[Xu \quad Xiaojie\_i \quad xiwang \quad Wang \quad Xiansheng \quad qu \quad ziji\_i.\]

Xu Miss hope Wang Mister marry.a.woman SELF

‘Miss Xu hopes that [Mr Wang will marry her.]’

3. Examples of grammaticalization in EMSEA languages

From handout 1 (p. 11): Properties of hidden vs. overt complexity:

**Overt complexity:** The structure of a language
(a) forces the speaker to overtly express certain grammatical categories (obligatoriness),
(b) provides a rich inventory of fine-grained grammatical categories.

**Hidden complexity:** The structure of a language
(a) does not force the speaker to overtly express grammatical categories that are part of its grammatical inventory (lack of obligatory categories)
(b) has multifunctional markers whose concrete meaning must be inferred from context
   (general context or linguistic/morphosyntactic context).

In handout 1 (p. 13), we have seen the example of Khmer បាន ba\n ‘come to have’. This verb was grammaticalized into three functions: (i) modality (possibility), (ii) past, (iii) truth/factuality. Khmer បាន ba\n ‘come to have’ is a case of multifunctionality. Thus, it is an instance of hidden complexity.

Other products of grammaticalization combine properties of overt and hidden complexity. Two of them will be discussed in this section:
3.1. Multiple analysis: The coexistence of older and newer forms in the case of ‘give’-verbs in Khmer

‘Give’-verbs can occur in various different constructions. Their concrete grammatical function depends on the construction in which they occur. The grammaticalization status of these verbs can be seen from the fact that each construction has its clearly determined word-order pattern with a specific slot for the grammaticalized ‘give’-verb.

(8) The functions of give verbs, presented in a grammaticalization cline:
   a. Verb > Coverb (preposition)
   b. Verb > Causative verb > adverbial subordinator > complementizer

(9) The constructions involved:

Coverb: [P NP], verb in the P position of PP (cf. (10))
Causative verb: [NP_{causer} CAUS (NP_{causee}) V NP ...] (cf. (11))
Adverbial subordinator: [Clause [ADVSUB Clause]] (cf. (12) and 13))
Complementizer: [Clause [COMPL Clause]] (cf. (14))

Some examples from Khmer:

(10) Coverb:

friend I come.to.have send letter COV:give I

‘My friend has sent a letter to me.’

(11) Causative verb (Bisang 1992: 440):

if CAUS I go be.involved.with I go
‘If you let me get involved with [him], I’ll go [there].’

(12) Adverbial subordinator marking purpose (Jacob 1968: 141):

I hard work PURP father I be.happy
‘I work hard to make my father happy/for my father to be happy.’

(13) Adverbial subordinator marking manner (Jacob 1968: 141):

Little.boy run go school MANNER be.quick
‘The little boy ran quickly to school.’

(14) Complementizer (Bisang 1992: 443):

she NEG want COMPL there.is matter what
‘She did not want that anything [bad] happens.’

While each construction has its individual word-order pattern, one and the same surface structure can be assigned to different constructions. This has mainly to do with the strength of “layering”
in terms of Hopper (1991), i.e., with the coexistence of the older and the newer functions with their constructions:

(15) Hopper & Traugott (2003: 49):

\[
A > \left\{ \begin{array}{c} B \\ A \end{array} \right\} > B
\]

The dominance of the \{B/A\}-situation in (15) makes it very hard to find monofunctional instances of B. Even some of the above examples are open to multiple analysis:

(12’) Coexistence of causative verb and adverbial subordinator (Jacob 1968: 141):

\[khŋom \ khom \ thvæt-\text{ka}(r) \ ?aoy \ ?oρukan khŋom \ səɾbajy-\text{cyt.}\]

I hardwork CAUS/ADVSUB father I be.happy

a. Causative verb: ‘I work hard and make father happy (by doing this).’

b. Adverbial subordinator (purpose): ‘I work hard for making father happy/ for my father to be happy.’

(14’) Coexistence of causative verb and complementizer (Bisang 1992: 443):

\[niŋη \ miŋn \ səŋ \ ?aoy \ miŋn \ ka: \ ?vvy.\]

she NEG want CAUS/ADVSUB there.is matter what

a. Causative verb: ‘She did not want to make something bad happen.’

b. Complementizer: ‘She did not want that anything [bad] happens.’

(16) Coexistence of coverb, causative verb and adverbial subordinator:

\[?oρukan \ səŋ \ phtəsh \ ?aoy \ kοm \ nɪu.\]

father build house give child live/stay

a. Coverb: ‘Father builds a house for his children to live in.’

b. Causative verb: ‘Father builds a house for making his children to live there.’

c. Adverbial subordinator: ‘Father builds a house with the purpose that his children live there.’

In most of its grammatical functions, ?aoy ‘give’ is part of a set of other markers with their specific meanings.

- **Prepositional function:** With verbs of transfer, there are two alternative strategies that highlight different aspects of the action of transfer:

  - Perspective of directionality: mɔʔk ‘come’, ʈuʔ ‘go’
    
    (10’’) Preposition/coverb:
    
    \[mʊt-\text{ṣəmlaŋ} \ khŋom \ bəm \ pʰŋəʔ \ səmbot \ mɔʔk \ khŋom\]
    
    friend I come.to.have send letter COV:give I
    
    ‘My friend has sent a letter to me.’ [in my direction]

  - Perspective of reaching destination: dəl ‘arrive, reach’:
    
    (10’’) Preposition/coverb:
    
    \[mʊt-\text{ṣəmlaŋ} \ khŋom \ bəm \ pʰŋəʔ \ səmbot \ dəl \ khŋom\]
    
    friend I come.to.have send letter COV:give I
    
    ‘My friend has sent a letter to me.’ [and it arrived at my place]
• **Causativity**: There are two alternatives of expressing causativity:
  - Morphological forms with certain verbs (lexical)
  - The verb ?aoy is limited to dynamic verbs. There is a semantically more general marker which is also used for stative verbs: thvː-?aoy [make-give] ‘CAUS’

• **Complementizer**: For many verbs, the complementizer is lexically determined. Some verbs allow more than one complementizer with different connotations. The verb prap ‘tell’ can have thaː (derived from a verb of saying) or ?aoy. With thaː, the embedded proposition is factual (17), with ?aoy it is not (18):

  (17) Khmer (Jacob 1968: 100):
  
  krùː prap kʰnom thaː kmeːŋ nùh prəlːŋ cəɾp.
  
  teacher tell 1.SG say boy DIST sit.for.examination pass
  
  ‘The teacher told me that the boy will pass the examination.’

  (18) Khmer (Bisang 1992: 442):
  
  kɔːt prap tɔːu kmeːŋ ?aoy yɔːk tɔːu cʊːn cɔmpɔːh
  
  3.SG tell to boy give take go give to
  
  nɛːk-ɛɾgɛ bɔndoːɭ.
  
  Miss Bondol
  
  ‘He told the boy that he take [the parcel] and deliver it to Miss Bondol.’

3.2. Different functions in different constructions:

**Numeral classifiers as markers of (in)definiteness**

This example illustrates the different functions of numeral classifiers in different functions:

(i) Individuation: [NUM-CL-N]

(ii) Referential status: [CL-N]

But even in (ii), the functional range is larger than in other languages. The classifier marks definiteness and indefiniteness in Sinitic languages.

3.2.1. **Basics on numeral classifiers [NUM-CL-N]**

The obligatory use of numeral classifiers with numerals:

(19) Chinese [NUM-CL-N]:

  a. sān fèng xīn
  
  three CL letter
  
  ‘three letters’

  b. *sān xīn
  
  three letter
  
  Intended meaning: ‘three letters’

The classifier is used to individuate the concept expressed by the noun and to make it accessible to counting (Bisang 1999).
3.2.2. The bare classifier phrase [CL-N] and (in)definiteness in Sinitic (Li & Bisang 2012)

The construction involved here only consists of the classifier (CL) and the noun: bare classifier phrase [CL-N].

The interpretation of [CL-N] in Putonghua, Wu Chinese\(^3\) and Cantonese in (i) the subject position (preverbal) and in (ii) the object position (postverbal):

(i) [CL+N] in subject position/preverbal position (Li & Bisang 2012: 338):

(20) A. Context: Where is the book?

   B. \textit{na ben shu, (\texttimes ge) xuesheng mai zou le.} [Putonghua]
   \begin{tabular}{llll}
   that & CL & book & CL & student & buy & away & PF \\
   \end{tabular}
   ‘That book, the student(s) has/have bought it.’

   C. \textit{b\text{\textring}{\texttimes}n cy, ky ia\texttimes san ma le ts\texttimes hi die.} [Wu]

   D. \textit{bun syu, go hoksaang maa\texttimes jo la.} [Cantonese]

   C and D: ‘The book, the student bought (it).’

(ii) [CL+N] in object position/postverbal position (Li & Bisang 2012: 338-339):

(21) a. \textit{wo mai le liang che.} [Putonghua]

   \begin{tabular}{lll}
   I & buy & PFV & CL & car \\
   \end{tabular}
   ‘I bought a car.’

   b. \textit{ny ma la bu ts\texttimes ow\texttimes hi.} [Wu]

   \begin{tabular}{lll}
   I & buy & PFV & CL & car \\
   \end{tabular}
   ‘I bought a car.’

   c. \textit{Keuih maa\texttimes zo gaa ce.} [Cantonese from Matthews & Yip 1994: 93]

   \begin{tabular}{lll}
   he & sell-PFV & CL & car \\
   \end{tabular}
   ‘He sold a car/ the car.’

Summary (Li & Bisang 2012: 340; I dropped the information on the topic position and the disposal position):

<table>
<thead>
<tr>
<th>[CL-N]</th>
<th>Subject</th>
<th>Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>Putonghua</td>
<td>---------</td>
<td>[-definite]</td>
</tr>
<tr>
<td>Wu Chinese</td>
<td>[+definite]</td>
<td>[-definite]</td>
</tr>
<tr>
<td>Cantonese</td>
<td>[+definite]</td>
<td>[+\pm definite]</td>
</tr>
</tbody>
</table>

Table 2: The functions of [CL-N]

---

\(^3\) Wu is a language situated between Putonghua and Cantonese. It is spoken in the Yangtze Delta area including Shanghai, Zhejiang Province and southern Jiangsu Province. The Wu data presented in this paper are all from the \textit{Fuyang} (富陽) dialect, which belongs to the \textit{Taihu} Lake group of the Northern Wu dialect. It has about 600,000 speakers in the \textit{Fuyang} city.
The interpretation of the (in)definiteness function of [CL-N] in Putonghua, Wu Chinese and Cantonese depends on:

- information structure
- word order and its association with (in)definiteness

1. Information structure:
   Topics are associated with high degrees of identifiability (Lambrecht 1994) and identifiability is associated with definiteness (Hawkins 1978, and many others):

   (22) Topic Acceptability Scale (Lambrecht 1994: 165):
   
   | active                                      | most acceptable |
   | accessible                                 |                 |
   | unused                                     |                 |
   | brand-new anchored                         |                 |
   | brand-new unanchored                       | least acceptable|

   Between topic and focus, there is a certain relationship of complementarity as far as definiteness is concerned. But, as is pointed out by Lambrecht (1994: 262), this is not necessarily the case:

   But this complementarity is only partial: while a topic constituent must have a referent, and while this referent must be identifiable and have a certain degree of pragmatic salience in discourse, a focus constituent is in principle free with respect to the question of identifiability and activation.

   In spite of this, the pragmatically inaccessible discourse referents from the lower part of hierarchy (22) are the most likely candidates for taking the function of focus even though a focus constituent is in principle free with respect to the question of identifiability and activation. Thus, focus NPs are usually indefinite, but definite ones are also possible in certain contexts (Lambrecht 1994: 262). In general, the focus position is less strictly associated with indefiniteness than the topic position is with definiteness.

2. Word order and (in)definiteness:
   As in many other languages, topics are preverbal in Chinese and there is a default postverbal position (Xu 2004) in Chinese. This correlation between information structure and (in)definiteness is more strongly grammaticalized into word order in Chinese:

   (23) Topic Subject disposal construction VERB Object

   |<-------------------------->|   |<-------------------------->|

   Definiteness                Indefiniteness (definiteness)

   On the grammaticalization from discourse-relevant structures to syntax, cf. Givón (1979) as discussed in handout 2 (p. 3):

   (24) Givón (1979: 209):

   Discourse > Syntax > Morphology > Morphophonemics > Zero

---

4 The use of the disposal construction (ba, kʰɔʔ, jeung) is a third factor, which will not be discussed here. For details, cf. Li & Bisang (2012).
3.2.3. Obligatoriness of the CL for marking (in)definiteness

The degree of obligatoriness of the classifier varies cross-linguistically. In general, it is not fully obligatory. Typically, a noun is introduced as indefinite and then marked as definite in its second and maybe third mention. Once it is familiar, only the bare noun will be used used.

This is illustrated by an example from Hmong. In this language, the classifier only marks definiteness/familiarity:

(24) Hmong (Mottin 1980):

```
Quav Tuv ho mus ntsib ib_pab qaib qus khiav khiav
Qua Tu then go meet/encounter one CL chicken wild run RED

1 los. Muaj ib tug plis caum los. Mas nws txawm come there.is one CL leopard chase come then he then
2 tib knee rau tus qaib qus nti ceev ceev. shoot.with.crossbow at CL chicken wild toss.to.and.fro quickly
3 Mas tus plis txawm dhia dhia los khawm nkaus thiab Then CL leopard then jump RED come l catch and
4 zaum tswag noj. Plis muab noj noj tag. Mas plis sit.down l eat leopard take eat RED complete then leopard
5 txhiaj li los hais yawg Quav Tuv tias: ...
6 then come say old Qua Tu QUOT
```

Qua Tuv came across a group of wild chicken which run towards him. There was [also] a leopard chasing after them. So, he shot at the wild chicken [and] they were tossing to and fro. Then, the leopard jumped towards [them], caught [them], sat down and ate [them]. After the leopard had eaten all of them, he came and said to old Qua Tuv: ...

4. Conclusions

1. On the independence of hidden vs. overt complexity in the two examples of § 3:

   (i) Grammaticalization of the verb ?aoy ‘give’ in Khmer:

   Hidden complexity: Multifunctionality
   Overt complexity: Preposition (P):
   There are other prepositions used with verbs of transfer, contributing different aspects of meaning:
   - verbs of movement: mĕk ‘come’, tṳ ‘go’
   - the verb dśl ‘arrive, reach’
   Causative verb: ?aoy is limited to dynamic verbs, thv?ːː?aoy can also occur with stative verbs.
Complementizer: 

?ao'y is not the only a complementizer. It still keeps some of its meaning as a full verb. For that reason, it contributes to the inventory of complementizers with there more fine-grained meaning:
- ?ao'y ‘give’: COMPL-function without assumption concerning factuality
- tha ‘say’: COMPL-function with the meaning of factuality

(ii) Grammaticalization of numeral classifiers as markers of (in)definiteness:

Hidden complexity: - Multifunctionality: marking of definiteness and indefiniteness
- Non-obligatoriness

Overt complexity: Fine-grainedness of the classificational semantics of classifiers:
There is a rich inventory of different classifiers whose selection is related to the semantics of the noun:
Just a few examples: 

zhī 只 ‘animals’
pǐ 匹 ‘horses and similar animals’
wéi 尾 ‘for fish’
kē 棵 ‘for tree-like objects’
tiáo 条 ‘for one-dimensional objects’
zhāng 张 ‘for two-dimensional objects’
běn 本 ‘for book-like objects’
zūn 尊 ‘Buddha statues, pieces of artillery’

The two examples clearly show the independence of hidden and overt complexity.

2. The processes of grammaticalization discussed in section 3 reflect both sides of maturation:

Economy-driven maturation  —> Hidden complexity
Implicitness-driven maturation  —> Overt complexity

In the next handout, we shall look at the two sides of maturation from the perspective of radical pro-drop.

References


