Language Change and Evolution

ESSLLI 17, 2005
Edinburgh

Course material
Regine Eckardt

1st Week, 9 - 10.30
Room EM 3.36
Part 1: Language in Time
Grammaticalization, Erosion, and Other Ageing Processes

The first part of this series of lectures is empirical in focus. We shall consider the question: What can happen to language over time? I present several cases studies in language change, mainly drawn from the European area for better accessibility. Clearly, no range of examples will ever be able to exhaust all processes of language change. Yet, the examples were chosen for specific reasons. They are either very prominent in the literature, thus offering salient material for further investigation and formal modelling. Or they exemplify language internal forces that are known to drive language change in a systematical manner. I try to avoid examples with known, local, accidental historical reasons.

We will pass the following range of topics:

- **Phonological reduction and its contrary**: It is a truism that speakers tend to reduce phonological material in quick speech, to avoid pronunciation costs. Independent words can shrink to become clitics and morphemes. Weak syllables can undergo vowel reduction and eventual loss. However, sometimes speakers also seem to enrich words with additional segments. We will see examples and ponder on possible motivations.

- **Regularisation and Suppletion**: Speakers of European Languages will be aware of the fact that much of inflectional morphology can depend on the morphological class of a given word (e.g. weak vs. strong verbs in English). On the one hand, irregular patterns tend to be regularised over time. On the other hand, however, there are also dynamic processes that lead to an increase of irregularity. Speakers can decide to use a word as member of an irregular class, counter to its actual origin. The process of suppletion (pushing in forms of one verb/noun into the paradigm of another) is another case where regularities get lost in history.

- **Building new tools**: Here, I will briefly draw attention to the general process of recruiting transparent constructions as a carrier for conventionalised non-transparent meanings. The process is often presented as one of the prototype instances of grammaticalization; the resulting constructions, at least in European languages, are typically classed as ‘analytic constructions’. We find an open source for new linguistic wraps for pieces of information that speakers feel the need to transport. Specific aspects of the general process will also be discussed in the following sections.

- **Bleaching**: I will use this term to refer to the loss of rhetorical or pragmatic information for a given entry. In terms of modern linguistics, these can be described as losses in presuppositions. We will review the case of negation where conventionalization and loss of rhetorical side messages are a substantial driving force in the process. (I will emphatically not use ‘bleaching’ as a synonym for ‘using a word in a more general sense, with a wider extension’.)

- **Other losses of pragmatic side messages**: This section will be devoted to an important and well-discussed change in the history of English, its development from an SOV to an SVO language. This change is taken as a paradigm case in much subsequent work and it is therefore worth knowing the details. I will argue that in spite of several attempts to capture the change in terms of formal syntax, it still looks as if the driving factors concern the conventions of information and discourse structuring. Hence, the change offers another instance for the losses of pragmatic side messaging.

- **System internal simplifications**: Finally, we will have a look at the development of the English Modal system. Lightfoot (1979) argues that the development consists of a gradual accumulation of exception rules and construction specific grammatical specifications which was then followed by a comparatively rapid reorganization. This reorganization resulted in the integration of modal verbs in the grammar of English, which follow a more and more systematic pattern of use and hence are presented as a simplification in comparison to the preceding language stage. Although Lightfoot’s presentation of the case has evoked...
controversial discussion, it still is a prime example of how language internal simplification processes could plausibly look like.

The overview should provide us with a variety of ageing processes of languages. Note that „ageing“ does not carry any evaluation here. I neither suggest that „older“ languages are „more mature“ languages, nor will we rely on notions like „degeneration“. I believe that most changes are adopted by speakers because the generations in question find a certain way to put things „better“ than all its possible alternatives. There may be many reasons, however, why one thing is better than another, and this section should also be understood as a warning against too narrow notions of improvement.

1. 1. Phonological Reduction and Epenthesis

Reduction:

(1) Nu loke euerich man toward himseluen
now look every man to himself
‘Now it is for every man to look to himself’

noghte > not
bletsian (inf.) > (to) bless
ða synfull-an sawle (acc.) > the sinful soul
hlaford > Lord

…

1.2. Regularisation and De-Regularisation of paradigms

Regularisation, example: Plural in English:

OE:  
cwen — cwene (queen/s)
schip — scipu (ship/s)
hund — hundas (dog/s)
sunu — suna (son/s)
eaga — eagan (eyes)
word — word (word/s)

c.1600 (Shakespeare): Choice between (e)s, (e)n
eyer, shooen, housen

Ongoing trend in foreign words

formulae — formulas
appendices — appendixes
data — datas

Remnants: man — men
child — children
sheep — sheep
goose — geese
(Brother — brethren)

Suppletion: Mixing of two paradigms:

go, goed > go, went
be—am, are—is—were
1.3. **Building-in new grammar**

The emergence of 'going to'

(2) a. Bill is going to go to college after all.
   b. Bill's gonna go to college after all.

(3) a. Bill's going to college after all.
   b. *Bill's gonna college after all.

Coexistence of older and newer ('grammatical') form. Morphophonological diversification.
Emergence of new form starting from one specific construction: purposive directional constructions.

(4) Bill is going to get some beer.
   a. Bill is going to the fridge in order to get some beer.
   b. There are predictable processes underway in the sequel of which Bill will soon get some beer.

Difference in meaning:
Older form: Movement component.
Newer form: Futurity component.

(5) Bill is *hastily* going (running off, sneaking away, ...) to get some beer.
(6) Bill is (*hastily) going to finish his talk at 16.15.

Reanalysis of underlying syntactic structure:

(7) [ Bill is going [ to get some beer ]]  
    [ Bill [is going to] get some beer ]

Phonological reduction witnesses the functional unity of *be-going-to* in analogy to other modals.
In the movement construction, these are not a functional unit. They can not be reduced (2.b).

In other respects, however, the *going to* future still functions as a transparent analytical tense form:

Existence of infinitival forms:

(8) Bill might be going to become a pianist.
    *Bill might ?will / willen/ ... to become a pianist.

Perception as a present tense aspectual rather than a future tense

(9) If interest rates are going to climb, we'll have to change our plans.
    *If interest rates will climb, we'll have to change our plans.

Observation: Grammaticalization often affects words that are rather general and abstract in meaning. (New future based on the verb *hop* seems unlikely).

Note: Synchronic diagnosis = Diachronic record of the change.

*Progressive*
Stage 1, Old English: Present Participle very similar to German:

(10) byrnende = burning / brennend  
lufiende = loving / liebend  
trymmende = strengthening / stärkend  
sellende = selling / verkaufend

Adjectival use, few predicative uses.

eall middangeard bi þonne on dæg byrnende  
all the earth is then by day burning  
(from Hogg, p.79)

*das schlafende Kind vs. ?das Kind ist schlafend*

Use in Old English does not cover all obligatory progressive contexts of PDE:

(11) os worold is on ofste and hit nacan ende  
this world is in haste and it is approaching its end.  
(Hogg, p.80)

Stage 2 (c. 1300): Due to phonological changes, several forms converge in the morphological form  
verb stem + *-ing*  
among which: present participle, nominalization(s)

(12) Reading is fun  
(Das) Lesen macht Spaß

(13) burning flames = brennende Flammen

New conventionalized uses (roughly since 1450):

(14) Sir James is at / on hunting  
Lady Mary is at / on writing letters

Compare variation of *am* progressive in Standard German and Ripuarian (Cologne area).

(15) Hans ist am Jagen  
John is at hunting  
Hans ist am Fliegenjagen  
John is at fly-hunting  
Hans ist Fliegen am Jagen (Cologne)  
John is flies at hunting  
Hans ist die Sau am Jagen  
John is the pig at hunting

Reduction of preposition:

(16) James is a-hunting  
The blacksmith was a-hammering and a-pounding

(Dialectal remnants in Appalachian English, Mid-east USA)

Emergence of modern progressive (obligatory in certain contexts; paradigmatic integration with simple tenses)
Perfect

Idealized development:
- I have used cars
  a. "I possess cars that have been used (by someone)"
  b. "I have used some cars"
- The subject of sentence and the agent of participle are equated
- Participle is reanalysed from an independent modifier of object into part of the main clause verb.

Full development attested for Latin + romance descendants (e.g. Drinka, Bridget 1997. The Evolution of Grammar: Evidence form Indo-European Perfects. in: Schmid, M. et al. (eds) Historical Linguistics 1997, pp. 117-133.). Old English, OHG etc. already show have-based perfects in the first written sources. This led sceptical voices to claim that the construction was adopted wholesale from Latin. (see Brinton, Laurel. 1988. The development of English Aspectual Systems. CUP.)

Careful statistical evaluation however proves that the construction was not common for all types of verbs. The use of perfect extended from verbs of intellectual possession (have an idea: expressed, have an objection: raised) to verbs of activity (have a foot: shoed) to intransitives and others (have: slept) (Carey, K. 1993. The grammaticalization of the Perfect in Old English. An account based on pragmatics and metaphor. In: Pagliuga (Hrsg.) Perspectives on Grammaticalization, 103-119. Amsterdam). These findings suggest that English still witnessed a development of the construction, which supports the hypothesis that an independent grammaticalization took place before and during OldEnglish.

'will' future, modal system

Functional elements in Syntax: 'that'.

Rudiments in OE suggest a deictic origin of complementizer 'that'. Typical examples are frequently of the form

(18) The king had that ensured, that the ceremony started in time.

(See Hopper+Traugott 1991/2003, chapter 7.3) Speculations are that the deictic 'that' originally was independent from the second clause:

The king had that ensured. That: The ceremony should start in time.

No clear examples are, yet, discussed in the literature.

Other source of complementizers: verbs with the meaning say. (See also Creole languages)

Discussion: Are discourse markers an instance of "grammaticalization"?

1.4. Bleaching: The Jespersen Cycle

In this section, we will investigate the origin of negation words in languages. I will argue that the development should be understood as a loss of presuppositions.

1.4.1. Negative Polarity Items:

Words, expressions, phrases that are only licensed in “negative” contexts
(19) Nobody knew **anything** about Apiz.
*Somebody knew **anything** about Apiz.
(20) Few humans **ever** saw Nessie.
*Many humans **ever** saw Godzilla

(21) Sue didn’t **bat an eyelid** when she saw Godzilla.
*Sue bat an eyelid when she saw Godzilla.

(22) Sabine hiel dem Angriff von Godzilla stand ohne mit der Wimper zu zucken
Sabine faced Godzilla’s attack without batting an eyelid
Sabine a recue l’attaque par Godzilla sans sourciller
(*sourcille* = eyebrow)

1.4.2. Negative Polarity Items are semi-transparent:

Negative polarity sensitivity is quasi-retained under translation

(23) einen Finger krumm machen
move a finger
lever le petit doigt

(24) einer Fliege etwas zuleide tun
faire mal à une mouche
hurt a fly

(25) keinen Pfifferling wert sein
not be worth a rap / a straw
ne pas valoir un pet de bique, un radis

The class of NPIs is semantically homogeneous

(26) **expressions of minimal value, minimal degree, minimal action**
mit der Wimper zucken
 einen Muckser tun
 roter Heller, Pfennig, Pfifferling
 einen Tropfen trinken
 einen Funken (Ehre, Anstand, ... ) haben
 ein Sterbenswörtchen

(27) **otherwise minimal on some scale**
 einen Hund auf die Straße jagen
 mit der Wimper zucken
 einer Fliege was zuleide tun
 die Miene verziehen
 den Finger regen

(28) **most general properties**
 eine Menschenseele, ein Schwein
 irgend-persone, rien
 any whatsoever

Etymological origin of NPIs
Deut = dutch copper coin, NPI since 17th ct.
Heller = copper coin from Schwäbisch Hall
grozs = Groschen

pas < Lat. passum ‘step’
point, mie = ‘crumb’, rien = ‘thing’, personne, goutte

noun > negative polarity item (most strong, one weak) > negation particle > …

1.4.3. NPI theories, the pragmatic account to Negative Polarity Sensitivity


NPIs evoke alternatives
Case A: Most-general-property NPIs
Alt( anybody ) = { P | P ⊂ [[ body ]] }
Alt ( Menschenseele ) = { P | P ⊂ [[ Mensch ]] }
   e.g.: { PASTOR, WIRT, BAUER, SCHUSTER, SCHNEIDER }

Case B: Minimal measure NPI (transparent case)
Alt( ein Tropfen ) = { P | P is a larger quantity }
   e.g.: { a drop, a mouthful, a glassful, a pitcher, a barrel }

NPIs stand in emphatic focus.
Emphatic focus states that the most striking, least expected, least probable among the alternatives at stake is true.

(32) emph.assert( S ):
   Assertion: [[ S ]]₀
   Presupposition: Against the given epistemic background of the utterance, and with respect to a salient scale (of likelihood, strikingness, salience):
   For all P ∈ [[ S ]]₁: [[ S ]]₀ <scale P

The (onto-) logical relations between NPI alternatives lead to logical relations between sentence alternatives.

(33) Bob drank a drop of beer
Alt(S) = { Bob drank a mouthful of beer, Bob drank a glass of beer, Bob drank a pitcherful of beer, Bob drank a barrelful of beer }

Logical ordering:
Bob drank a drop of beer
Bob drank a mouthful of beer
Bob drank a glass of beer
Bob drank a pitcherful of beer
Bob drank a barrelful of beer

**Logically weakest** ⇔ **least striking, most likely**

**Logically strongest** ⇔ **most striking, least likely**

**Pragmatic NPI licensing:** An NPI can only be acceptable iff the sentence in which it occurs is the logically strongest among its alternatives.

(34)  
Gary did not drink a drop of beer.  
Alt(S) = \{Gary did not drink a mouthful of beer, Gary did not drink a glass of beer, Gary did not drink a pitcherful of beer, Gary did not drink a barrelful of beer\}

Logical ordering:

- Gary did not drink a drop of beer
- Gary did not drink a mouthful of beer
- Gary did not drink a glass of beer
- Gary did not drink a pitcherful of beer
- Gary did not drink a barrelful of beer

Result: “a drop” in emphatic focus is acceptable, because the resulting sentence is the logically strongest one.

1.4.4. From NPI to plain negation

According to the pragmatic account of NPI licensing, the core difference between NPIs and other words and constructions consists in a special lexical entry concerning their rhetorical function. They carry a requirement that

- they have to be used as a rhetoric device
- they have to be used in emphatic focus
- they evoke focus alternatives, sometimes in a non-standard way (= most general property NPIs)

The patterns of use follow from these requirements.
Language change sets in where speakers of following generations fail to understand the obligatory rhetoric quality of a locution. Disregarding the alternatives at stake, plus the presuppositions about likelihoods of different alternatives, the sentence meaning comes down to a plain negation. Different kinds of development can set in.

In French, we know that the earlier Negative Polarity Items *rien*, *pas*, *mie*, *goutte*, *personne* and *point* lost their rhetorical quality between Old and Middle French, *point* being the youngest item in question. In the course of that development, all uses in NPI contexts apart from negation *ne* are lost, apart from isolated conventionalized locutions. However, speaker communities can retain a stylistic sense for some negation being ‘more emphatic’ than plain negation. French *point* offers an example. It showed systematic uses as an NPI till Middle French (c. 1500, xx), which is about 500 years later than other items in the series. Teachers of French comment on the difference between *point* and *pas* as the former being ‘stronger’ as a negation, signalling more emotional commitment of the speaker. The full pattern of emphatic focus use has, however, become obsolete in Modern French.
The outline of this development was already pointed out by Otto Jespersen (1917xx) and has become known as the „jespersen cycle“. Its instantiation in English is as follows:

Old English possessed a simple sentence negation _ne_ which was used preverbal. It also could show Negative Concord if necessary.

Consulting an etymological dictionary, we learn that negation in Modern English derives from agglutination of _ne_ with other material, mostly indefinites. _ne + æfre_ = ‘never’, _ne + a-wiht_ ‘not a tiny-person’ => _na-wiht, noht_ ‘nothing’ => PDE _not_

The road from these indefinites via emphatic negation to simple negation follows the lead predicted by the pragmatic account of polarity sensitivity. I will use _not_ as an illustration. _ne + ough_ ‘not the least thing’ plausibly went through the following stages:

(a) transparent use in appropriate negative sentences like in

‘Tom _ne_ found _ne_ the least thing’

Note that the ‘true’ negation is located preverbally while the negation _ne_ before the noun _wiht_, which will agglutinate into the new negation „not“, is used due to negative concord.

(b) conventionalized use as emphatic negation

‘Tom _ne_ slept _not-the-least-thing_’

The former noun _wiht_ is used as a minimal measure NPI beyond its former literal meaning. It is still used as an emphatic negation. _Wiht_ can also be found in other NPI licensing contexts at that stage (a fact that is reported in footnotes or late entries in dictionaries of Old English, as well as Old High German)

(c) loss of emphasis, simultaneous loss of preverbal negation.

‘Tom slept not’

The former NPI has been acquired by later language learners without a lexical requirement to be used as carrier of emphatic focus. While the literal content „negation“ has been retained, the pragmatic side messages have been lost: _bleaching_. (The subsequent development of do/modal support for negation is seen as an independent step; see German, French.)

1.5. English: From SOV to SVO

1.5.1. Data and crucial data.


Verb second: Word order in Old English varies, like in many modern Germanic languages. Embedded sentences show a trend towards SOV while main clauses are mostly verb-second. But even then, verb-second sentences are not the same as in modern English.

Modern English: Subject - Verb - Object
Old English: XP - Verb - Object - (infinite parts of the verb).

The older stage is characterised by V2 movement. The verb moves from V0 to I or C, the preverbal constituent likewise arrives by movement. (For details that distinguish between the two analyses, see Fischer et. al., chap. 4) The moved element characterized by its discourse status; Fischer et al. characterize it as „topical“ (p. 136). Loss of V2 is hence one step that turns English from a discourse configurational language (word order driven by information structure) to a language with fixed word order (word order reflects argument structure).
**O-V word order:** Several modern Germanic languages (e.g. Dutch, German) show consistent correlation between main clauses and V2 pattern and subordinate clauses in V2nd pattern. Old English likewise is an example for this distinction, even though the correlation is a statistical trend rather than a fully coherent rule. The following facts exemplify the grammar of Old English (all taken from Fischer et al, Chap 5; consult for original sources.)

Position of finite verb in part dependent on type of clause, although OV and VO both occur even in subordinate clauses:

(35) āet he hwelc gerisenlic wundor wyrceanmæge
    that he some suitable miracle perform can

(36) āet he nolde niman mancyn neadunga of ðam deofle
    that he not-would take mankind forcibly from the devil

Negative adverb *ne* is laways on the immediate left of the finite verb (proclitic)

(37) ðæt he na siðan geboren ne wurde
    that he never afterward born not would-be

(38) ne sende se deofol ða fyr of heofenum
    not sent the devil then fire from heaven
    þeah þe hit ufan come
    though that it from-above came

Verb particles: often to the immediate left of the lexical verb (note: verb is final here)

(39) gif hio ne bið hrædlīc aweg adrifen
    if she not is quickly away driven

(40) ðæt hie mit þæm þæt folc ut alloccoden
    that they with that the people out enticed

Other particles: sometimes higher in the clause; but then likely to be part of PP:

(41) þæah þu sie up ofer ðine mæð æhæfen
    though you are up over your condition raised

Stranded prepositions: nearly always left-adjacent to verb. Very rare examples with preposition right-adjacent to a non-fronted verb:

(42) þa wæs hiora an se Apollonius þe we ær ymb spræcon
    then was of-them onethe Apollonius that we before about spoke

(43) þæt is þæt uplice rice, þe he sylf wunaþ
    that is the heavenly kingdom that he self lives
    on mid eallum his halgum a butan ende
    in with all his saints forever without end

(Pronouns ... xx)

Adverbs: Usually to the left of any non-fronted verb

(44) Simon ... getigde ænne ormærne ryððan innan ðam geate
Simon tied a huge mastiff inside the gate. 
ære he ferlice hine abitan sceolde 
so-that he suddenly him bite  might

Very rare examples where adverb follows an infinitive.
Nominal objects and PPs may occur anywhere between subject and non-finite verb. 
Likewise, bare infinitival complements can occur preverbal.

(45)  He mæg ða synfullan sawle durh his gife geliffæstan  
he can the sinful soul through his gift endow-with-life

(46)  þæt he ure sawle fram synna fagnyssum gehælan mæge 
that he our soul from of-sins ulcers heal can

(47) Da se Wisdom þa þis fitte asungen hæfde,  
when the wisdom then this poem sung had
þa ongon he eft seggan spell  
then began he again tell story

All these elements can, however, likewise occur to the right of a non-fronted verb. 
Clausal objects always occur to the right of all verbs.

1.5.2. The underlying structure(s). Some speculations in X’-syntax

Talking about SVO and SOV word orders, we seem to suggest that the only change to be traced 
consists in a difference between two word orders. 
Some authors also, simplifying, point out the almost complete loss of inflection endings between 
OE and EModE and the concomittant difficulties to reconstruct a constituent's abstract case in the 
sentence—unless speakers could rely on strict word order. This observation would suggest a move 
from free word order to strict word order.

In reality, the development is much less perspicuous. If we rely on word order patterns know from 
other Germanic languages, and if we believe that grammatical facts and rules of present day 
Dutch, German, Icelandic, Swedish etc. can set the range for possible and likely grammars of Old 
English, we get the following spectrum of rules and structures involved.

(a) Head-initial versus head-final phrases, in particular VP and IP. 
If verbs or verb forms can be found with a certain reliability at the end of clauses, we should 
expect that the underlying position of V in VP is at the right end (= final) and that, if we have 
reason to believe that finite verbs undergo movement to Infl, I is likewise final in IP.

(48) 

This is more or less what is assumed for modern Dutch and German.
If the heads of IP and VP are initial, we assume underlying syntactic structures like the following, as is the case in modern Scandinavian languages.

(49)

\[
\begin{array}{c}
\text{SpecI} \\
I \\
\text{SpecV} \\
V \\
\text{Compl.}
\end{array}
\]

Looking at examples from Old and Middle English, we will expect that their underlying syntactic structure lies somewhere in the range of these possibilities, and we will accept the same key data as support for one or the other analysis as we do for modern languages.

(b) Verb-Second movement (V2)

In Old English as well as modern German and Dutch, we find main clauses where the verb is reliably in second position. The first position can, however, be occupied by different constituent types. In modern Dutch and German, we have a full range of data that can be tested, and the connection between "Topicalization" (some constituent being moved into initial position) and discourse structure is well-understood. Without discussing all details here, it seems clear that a constituent can be fronted if it is particularly in focus or if it is the topic of the ongoing discourse i.e. particularly 'old'.

It is assumed that V2 moves the finite verb to C, and that the focussed-or-topicalized constituent occupies SpecCP. Evidence in favour of this assumption mainly comes from systematic asymmetries between subordinate and main clauses: V2 in the sense described here is, in modern languages, a strict main-clause phenomenon.

It is assumed that Old English likewise possessed V2 movement, in view of the fact that many clauses start with adverbial elements. This conforms with the topic-function of SpecCP-movement in modern Germanic languages: Reference time or reference place are often the topical anchor which relates a new part of the story to preceding context.

(c) V-to-I movement

Lightfoot, 1999 points out that the special status of the inflection head in modern English in part derives from the fact that only modals and auxiliaries can be found there. This assumption takes care of the well-known facts about do-support for negated sentences, yes-no questions and wh-questions in terms of X'-structure. While V2-movement needs V-to-I movement as a necessary precondition (if we trust the principles of X'-theory according to which heads can only be moved via intermediary "nearby" landing sites), we could also conceive of a grammar without V2-movement but V-to-I movement. We'd expect a language with uniform Subject-initial clauses but without do-support in questions and negations.

The data in the history of English do not support a stage with V-to-I movement and without V2. However, Lightfoot correctly stresses that a consistent account of the history of English in terms of language acquisition not only requires a story for the loss of V2 movement, but also a story for
the loss of V-to-I movement, given that children at earlier stages acquired a grammar that possessed both movement rules.

Unfortunately, the formal accounts of the history of English syntax, as far as I can tell, remain tacit about the syntactic nature of such movements as heavy NP shift or scrambling as a putative option in OE. These, as well as other "stylistic" movements are notoriously hard to fit easily into X'-syntax. Yet, discourse and processing driven word order variants are undeniable parts of grammars of older and modern Germanic languages and it is evident that variation at that end plays a crucial role in the radical grammatical change that took place between 900 and today.

1.5.3. Ways of change

Discourse and processing reasons

Heavy NP shift or VO

In modern SOV languages, we know that certain kinds of arguments, or parts of arguments, tend to be postposed to the right of the verb. Informally, factors like short term memory and dispreferred endocentric embeddings seem to play a role. The following contrasts hold in Modern German:

(50) (weil) John [den Mann, der einen Hut trägt], gesehen hat.
    c. ?? (weil) John [den Mann, der gestern mit Maria dreimal die ganze Stadt auf dem Motorrad umrundet hat], gesehen hat.

(51) a. ?(weil) John, dass es regnen wird, vorhergesagt hat.
    b. (weil) John vorhergesagt hat, daß es regnen wird.
    c. *(weil) John, daß es regnet, sieht.
    d. (weil) John sieht, daß es regnet.

Topic-first or subject-first

In V2 languages, the initial constituent can either be in focus—a constellation which is usually very clear in utterance context—or else be the topic. Without any deeper investigation into the nature of "topichood", let us take it for granted that there is a universal statistical correlation between subject arguments of verbs and likely topics in sentences that contain these verbs. (Simply put, it appears to be a universal in communication that a sentence like, for example, 'A saw B' is more often a sentence about A than a sentence about B.)

The majority of other pre-verbal consituents in V2 sentences of OE are adverbs of time or place like tha or thonne (= 'there' and 'then').

Without being able to offer precise counts at this point, it seems likely that a small range of conventionalised sentence-initial constituent types is likely to obscure the information structural motivations for fronting these XPs.

Discourse driven, stylistic or grammar driven word orders

I would like to mention that not only processing reasons can license postposing material to the right of the infinite verb part in modern Germanic languages. Adjuncts of the verb (temporal, local, instrumental, etc.) can easily be postponed and receive the quality of an afterthought. Excessive use of patterns like in (52) might contribute to obscuring the original head position in VP and IP.

(52) (weil) Susanne den Mann noch gesehen hat, gestern.
    (weil) Inga das Buch gelesen hat, im Bus.
    (weil) Anna die Fenster schon gewischt hat, mit einem Tuch.
These options exist alongside the often-quoted freedom of word order in verse. While they once more require specific kinds of utterance situations and preceding discourse, it can easily be seen that they exhibit word order patterns that would be puzzling for the historical syntactician as well as for contemporary speakers who fail to understand the discourse licensing factors.

"Pronouns tend to stay left"
In modern Germanic languages with 'scrambling', movement in the so-called middlefield, we observe that pronouns, unless they are in narrow focus, are required to stand as far to the left as possible. The following examples illustrate parts of this pattern.

(53) (weil) Hans sie dem Freund geschenkt hat.
    (weil) *Hans dem Freund sie geschenkt hat.
(54) (weil) Hans ihm das Auto gezeigt hat.
    (weil) *Hans das Auto ihm gezeigt hat.
(55) (weil) sie Hans dem Freund geschenkt hat.
    (weil) ihm Hans das Auto gezeigt hat.
    (weil) er die Vase dem Freund geschenkt hat.
(56) (weil) *die Vase er dem Freund geschenkt hat.
    (weil) *dem Freund er die Vase geschenkt hat.
    etc.

Such observations cohere with the facts of OE insofar as the pronouns (together with quantificational and negative DPs) are found before the verb till very late (15th century).

Quantificational and negative NPs tend to stand before the Vfin, even at at time when the overwhelming majority of sentences conformed to modern SVO (van der Wurff). This is a pattern that does not follow straightforwardly from anything that we know about the information structural status of these types of DPs. Which is little enough, anyway.

Some hypotheses:

=> where the older grammar could analyse an "early verb" (particularly in subordinate clauses, or if the verb was infinite) as "early" because everything else was shifted to the right, the newer grammar could analyse these sentences as the underlying base order.

=> where a light object (pronouns) was placed before the verb, the older grammar could conceive this as the underlying base position, whereas a more modern (intermediary?) grammar could analyse this as a discourse driven aberration from the underlying SVO order.

Even the recent Fischer et al. 2001 do not spell out any such intermediary grammars in detail which systematically are SVO and possess additional discourse-driven options to provide for OV order in well-defined contexts (i.e. a grammar with well-defined interfaces between discourse and syntax). We know from the investigation of scrambling in Dutch/German that this is an extremely subtle task, even if one has access to living speakers...

Language contact
Pintzuk and Kroch (quoted from Fischer et al., 2000) diagnose a diglossic situation in England between 1200 and 1400. They offer evidence that the intrusion by Scandinavians from the north had led to a head-initial dialect with V2 movement. Pintzuk argues that after centuries of factual diglossic speakers, the situation resolved in favour of the modern instantiation.

Features strong and weak
Fischer et al. elaborate an alternative proposal that is couched in terms of Chomsky's minimalist program. The main feature of the proposal, seen in terms of language change, consists in the
assumption that all languages have underlying SVO word order. This order can be changed by movement which takes place to certain functional positions in order to regulate feature assignments (feature checking). The relevant functional positions are located such that ouvert movement will lead to the word order facts of other languages, notably SOV.

Movement can take place before or after spell-out, i.e. either at a stage in the derivation that leads to visible movement, or at a stage that corresponds to covert movement. Which kind of movement takes place depends on the “strength” of the features that regulate the movement. So-called strong features are responsible for ouvert movement while weak features lead to hidden movement.

This approach, although in need for clarification in the details (as the authors readily admit) would have the advantage that a very radical change in grammar—word order facts—can be mimicked by a seemingly softer change in grammar—the change of strength in some features. The choreography of syntax in the Chomskyan tradition suggests that changes in features are less challenge to UG, and less a break between the parameters of adult language and the following children’s generation than changes in headedness of VP and IP. Yet, this intuition should be reconsidered in the light of the new setup in the minimalist program. If features are so powerful as to determine word order facts in a given language, and do so very reliably and unchangeably under normal First Language Acquisition, the assumption that speakers and learners can maintain misunderstandings and support variation at that point appears to be a severe challenge to known facts about the acquisition of word order, and in particular very early and reliable preferences for SVO vs. SOV by children of SVO or SOV speaking parent generations, respectively (see for instance the work by Meisel on comparative studies and studies on the bi-lingual acquisition of French and German).

Finally, Fischer et al. admit that speaker variation and preference for weak and strong features should not be random, but driven by soft communicative factors like discourse and processing preferences. This however means that the approach is in need for exactly the kind of supplementation theory that the authors criticise harshly as being not faithful to the data, not detailed enough, and of no explanatory power.

Therefore I submit that, elegant as the approach may be at first sight, it just seems to rearrange the same old questions in a new manner rather than answering them.

1.5.4. Acquisition based viewpoints

Cues and grammars

Lightfoot points out that children do not acquire grammars in a simple pattern-matching manner. They do not attempt to copy larger and larger parts of the target language until they reach full competence.

They rather seem to filter out certain elementary structural properties and build learner's grammars on basis of these properties. Such learner's grammars reflect some underlying features of the target language without necessarily producing "correct" sentences of the target language. (Example: consistent early use of verb-final infinitives by German children: Apfel essen, Papa schlafen, Buch gucken etc.)

Against this background, Lightfoot proposes that the primary linguistic data (PLD) need to offer certain cues, data that are crucial for the language learner to understand some elementary feature of the target language. He likewise proposes that cues can cause the child to give up some current grammar and move on to a more complex grammar. He finally assumes that such moves from simpler to more complex grammars need a certain frequency of relevant cues. Children will not change their grammars on the basis of isolated examples which might be accidental errors.
Lightfoot discusses two cues in particular, the V2 cue and the V-to-I cue. He argues that sentences with arbitrary non-subject initial constituents provide cues for V2 movement to C (and one further constituent to SpecC, see above). For this particular cue, real counts are available for V2 languages and the transition stage of Middle English. Lightfoot reports that counts of acquisition data PLD for modern Dutch contain about 30% cues for V2 movement. Textual samples in ME, in contrast, contain only about 10% cue sentences for V2 movement. He therefore hypothesises that

the crucial threshold for a cue to show an effect on the learner's grammar is somewhere below 30% and above 10% (given that V2 was actually lost in ME).

The V-to-I cue is provided by V-movement in questions and negated sentences and seemed to be likewise lost in ME due to the rising proportion of sentences with do-support (compare development of modal verbs).

Lightfoot does not propose any cues for head-final vs. head-initial VP and IP. Therefore his work falls short of a full account as the essential shift of the verb from sentence final to sentence second position remains untreated. He refers to work by O. Fischer in the feature based model (minimalist program). However, the scepticism against Fischer et al.'s model carries over to any subsequent theory that rests on it.

One feature in Lightfoot's model should be highlighted: His view allows for purely language internal changes. Particularly, as soon as certain cues occur with low frequency in the PLD we expect that children cease to change some intermediate learner's grammar in the relevant way in the direction of the older parental grammar (e.g. fail to allow for V2 movement).

1.6. Structural Simplification: The Modal System

Development towards a language-internal simplification of a range of constructions fraught with exceptions. Lightfoot, 1979, on the emergence of modals.

Modal Verbs in Modern English: must, shall, will, can, may, ought, might, need
Pre-Modals: The predecessors of these verbs which did not yet share the relevant grammatical properties, verbs that would later on develop into a modal.

Stage 1: Pre-Modals as fully lexical verbs

(57) scullan > shall / should
willan > will / would
magan > may / might
cunnan > can / could
motan > must

full inflection paradigms, including infinite forms, gerunds, etc.

Pre-Modals are serialisable (all examples see Lightfoot 1979: chap. 4):

(58) I shall not konne answere
    I will not be able to answer

(59) Cunnyng no recour in so streit a neede
    Knowing no recourse in so desperate a need.
(60) if we had mought conuenient come together
     If we had been able to meet conveniently

(61) if he had wolde
     if he had wanted to

Negation patterns with other verbs:

(62) zif þe wollnot to haue mercy of God
     If you don’t want God’s mercy

(63) Thy godfadirs wyff thow shalt not take
     you shall not take your godfather’s wife

(64) A blynde man kan nat juggen wel in hewis
     a blind man can not judge colours well

(65) He ne held it noght
     He did not hold it

(66) My wyfe rose nott
     My wife did not get up

Question formation patterns with other verbs:

(67) Wilt thow ony thinge with hym?
     Do your want him for anything?

(68) Truwast þu nu þe selfum and þinum geferum bet þonne þam apostolum … ?
     Do your trust yourself now and your companions better than the apostles … ?

Allow for NP objects:

(69) she koude muche of wandrynge by the weye
     she knew much of wandering by the way

(70) seþ sculde him undred denera
     he who owed him a hundred denarii

(71) ac him sealde þa mihte se þe mæg ealle þing
     but he who can do all things gave him the power

sentence final in SOV sentences (like other verbs)

But: Pre-Modals are „preterite presents“ which sets them (and other verbs in the same class) apart from the common inflection paradigms.

2: Slow developments

Nominal arguments recede. E.g. last examples in OED: shall 1425, can 1649, may 1470

Inflection paradigm gets more and more isolated:
• other earlier, non-pre-modal verbs in this class get extinct
• distinction between past tense forms and subjunctive forms get blurred (compare: *konnte — konnte and *konnte — konnte)
• loss of the transparent past-present relation
  Kim will visit Rome — Kim would visit Rome (possibility; past future)
  Kim can play tennis — (In former years,) Kim could play tennis / Kim could play tennis, (if she wishes to)

The to infinitive gets established for more and more clausal complements. Yet, complement clauses of pre-modals are systematically spared.

At the same time: development of periphrastic do constructions. From ca. 1400, presumably from causative constructions like the following:

(72) He dude writes sende
    = 'he let (someone) send letters'

(73) he dude strepe this Maide naked
    = 'he let / ?did stripe the maid naked

While an initial emphatic use is not unlikely, the details of the development of periphrastic do have not been elucidated to full satisfaction.

Stage 3: Radical Reanalysis

Establishing a new class of modal verbs, patterning with the auxiliaries.

Transparency Principle: In language acquisition, learners prefer a simpler structural analysis + lexicon for a given class of constructions over an analysis that rests on extra assumptions, exception rules etc.

Special inflectional status: Modals were known to be different in inflection. Modern reflex: no -s in 3rd person singular

No infinitive forms: Modal verbs fail to have an infinitive form
  *to must, *to can, *to will, ...
Last examples in texts around 1550. Last gerund forms (-ing) disappear around 1550. Last analytic perfect have +V-en (habe cunnen) disappear around 1530.

At the same time, quasi modals be going to, be able to, be obliged to are conventionlized. They supply full inflection paradigms.

• Modals can not be serialised:
  (74) *John must can read „War and Peace“
       *Susan will must read „War and Peace“
  (75) John must be able to read „War and Peace“
       Susan will have to read „War and Peace“.

Last examples of serialised modals disappear around 1532.

• Modals stand in for do-support in questions
  (76) Does Susan know Duffy Duck?
       Should Susan know Duffy Duck?
       *Does Susan shall know Duffy Duck?
(77)  Whom does Susan invite?
    Whom will Susan invite?
  *Whom does Susan will invite?
Inversion rule: „Invert, and invert only modals and auxiliaries“ finds first application in 16th century. Old polar question with lexical verb in initial position die out in 17th century.

* Modals stand in for do-support in negations
(78)  Susan does not know Duffy Duck.
    Susan should not know Duffy Duck.
    *Susan does not shall know Duffy Duck
Negation always after first auxiliary or modal in the sentence, spread of patterns since c.1500

* Modals stand in for do in emphatic constructions
(79)  Susan DOES love Beethoven.
    Susan SHOULD listen to Beethoven.
    *Susan DOES should listen to Beethoven.

* Modals do not take full NP object complements. (see above)

Lightfoots analysis was received with scepticism: Doubts in the speed of the „radical reanalysis“ part of the picture. Doubts in the sudden emergence of all those constructions that are typical for present day modals. Doubts in the emergence of all constructions for all modals at the same time.

* Nevertheless, the example can serve as an idealized scenario of what may happen if new generations of speakers perceive a simpler syntactic analysis for a whole range of constructions as more plausible than an older, more complex one.
Part 2: Language from Zero.
Creolization, Koinezation, and How Languages Get Less Complex

2.1. Pidgins and Creoles

(John Holm. 2000. An Introduction to pidgins and creoles. Cambridge Textbooks in Linguistics.)

**Jargon**: Interlocutors agree on content words. Expression of syntactic functions is left to context or decided in an ad hoc fashion. No conventionalization in a speaker community.

**Pidgin**: Communication mainly on the basis of content words, often with a restricted range of topics (trade, plantation). Communication between different language families (otherwise: koinezation: mixing of varieties). Not a language of ethnic reference. (Critical cases are known, like e.g. Tok Pisin which is not conceived as the language of ethnic reference of any of the interacting groups.)

Typical constellation: Social distance between speaker groups. Balanced situations are known (Russenorsk) but more typically, there is

- a socially dominant group = superstrate language = lexifier language
- a socially inferior group = substrate language = role controversial?

Different situations of contact lead to different further developments.

Examples:

```
Then I stay Rabaul I was in workgroup I
katim kopra. naw wonfəb master ḅɔŋ kampani əm
Cut kopra. Then a white man from company he
i-kəŋə mi mi kək ḅɔŋ əm gen. master king, mi
take me I cook for him again. Master King, I
stap. naw ol master i-kək əm.
stay Then all white men were playing football. They-kick him.
naw leg ḅɔŋ əm i-swelap.
Then leg of him swell up
```


Hawaiian Pidgins, spoken by Japanese (a), Korean (b) and Filipino (c) immigrants (from Bickerton, D. 1990. Language and Species. Uchicago press: 120). Italicised words are of Hawaiian origin.

(a) Ifu laik meiki, mo beta make time, mani no kaen hapai
   If you like make, more better die time, money no can carry
   If you want to build (a temple), you should do it just before your die—you can’t take it with you

(b) Aena tu macha churen, samawl churen, haus mani pei
   And too much children, small children, house money pay
   And I had too many children, small children, and I had to pay the rent

(c) Luna, bu hapai? Hapai awl, hemo awl
   Foreman, who carry? Carry all, cut all
'Who’ll carry it, boss? Everyone will cut it and everyone will carry it'

Some grammatical elements are present: conditional *ifu*, negative *no*, conjunction *aena*, auxiliary *kaen*, quantifier *tu macha*, question pronoun *hu*, and quasi modal *mo beta* ('should'). Missing: Articles, prepositions, complementizers, tense, aspect.

**Creole.** All creoles have a pidgin/jargon in their ancestry. Creoles are, however, native languages within a community. Main social precondition: Removal of community of ethnic/linguistic roots.

1. **Plantation creoles:** Removal from ethnic origin, systematic mixing of linguistic groups
2. **Fort Creoles:** Mixing of superstrate culture in bilingual, pidgin-based marriages.


(d) They wen go up there early in the morning e go plant
   'They went up there early in the morning in order to plant (crops)'

(e) I gotta go hore one carpenter e go fix the form
   'I have to hire a carpenter to fix the form'

→ Regular grammatical and thematic relations, tense, aspect; use of articles, clause embedding.

→ Word order variations related to function; ungrammatical word orders exist:

(f) i. Some guys they get different belief
   'Some guys have different beliefs'
   ii. They get different belief, some guys
       right movement, some guys, de-emphasized
   iii. Different belief some guys they get
        fronting of 'different belief', emphasis
   iv. Different belief they get, some guys
       emphasis of 'diff. belief', de-emphasis of 'some guys'
   v. *Some guys, different belief they get.
   vi. They get, some guys, different belief.

→ Embedded structure and adjunction:

(g) [The guy gon' lay the vinyl] been quote me price
   'The guy who was going to lay the vinyl had quoted me a price'

**When does a pidgin start to be a creole?**

Linguistic indications:

*Extension of the lexicon* beyond a clear range of economic contact issues.

*Grammar:* Extension of the tense-mood-aspect system. Word order regulation, fixed patterns of assertive sentences, questions, subordination, relative clause formation, complementizers, serial verbs (for the minimal grammatical tool-kit, see Bickerton, below)

*Phonology:* Raise of phonological processes of regular language like assimilation.

*Social* indications: Adoption of language as a language of ethnic reference.

*Acquisition as a first language.*

**How does a pidgin turn into a creole?**

Prominent hypotheses:

- early/ abrupt creolization in first language acquisition. (Bickerton)
- slow/gradual creolization, mainly in adult speaking communities (e.g. Mufwene)

... more in section 2.2.

**When does a creole stop being one?**
Continued contact between language communities leads to a continuum of varieties with different degrees of overlap with lexifier language:

basilect — mesolect —acrolect

Yet, Holm believes in the notion of "loss of creole features" and the development of "post creoles" with seems to indicate that a variety can be

- different from the lexifier superstrate language and yet
- no longer a creole language. (AAVE, Brazilian Vernacular Portugese)

Appendix: Cases of mixing

Mbugu = Kushitic lexicon and full Bantu grammar
Anglo-Romani = English syntax, phonology and grammar plus Romani content words for in-group communication
Michif = mixture of Cree verb phrases + French noun phrases. (see Holm, p. 11) Spoken in Turtle Mountain Indian Reservation, North Dakota. Is generally taken to be product of volitional mixing to define culture of mixed people (Fr. métis = michif?)

\[ Nkii-cihtaan \quad \text{I-PAST-go} \quad \text{dans la ligne} \quad \text{to-theFEM} \quad \text{state-line} \]

Holm (2000) takes six creoles as the starting point for his discussion and offers an illustrative survey over the recent history of the respective speaker communities.

Map of Creoles

2.2. The emergence of a Creole

Nurture Hypothesis: Adult developments. Piecemeal accumulation of new grammatical forms as in common grammaticalization. Main source of grammar: Substrate and Superstrate, copying of known constructions into the newly emerging language.

Nature Hypothesis: Grammatical patterns are spotted and conventionalized in L1 acquisition. The main subgroup of speakers who shape the emerging language is the group of children up to ca. 10y. The grammatical patterns are not „copied“ from any full language but emerge in the I-language of children in acquisition with impoverished stimuli.

Possible variants of the Nature Hypothesis:

1. Gaps in the input grammar are filled according to default parameter settings in UG (early Bickerton)
2. Gaps in the input grammar are filled by conventionalising and structuring accidental traits in the input.
3. Gaps in the input are filled by turning the impoverished input into some grammar in line with the default expectations in UG („no trigger input will ever be met that indicates that the simplest analysis of a sentence was not the analysis of the target language“).

Arguments in favour of the Nature Hypothesis:

for a moderate variant of the bioprogram:

- Creoles can arise under different circumstances. Slow and gradual emergence ("fort creole") differs qualitatively from quick and radical emergence ("plantation creole").
- Creoles are not always a conservative extension of previously existing pidgins (see Bickerton on Hawaiian Creole English)
- In the initial stage, there is loss and retention of forms of the superstrate language (depending on: historical circumstances, allomorphy vs. none, conformity to canonical syllable structure, markedness, and other factors, in part poorly understood)
- In a second phase, there is reconstitution of the necessary prerequisites of a full natural language.
- Ideally, reconstitution of the absolutely necessary elements of grammar should take place in 1 generation.

Minimal elements of a full natural language (Bickerton 1988: 278)

Grammatical functions that will be reconstituted:

1. articles (note: languages that do not have a. count as more marked/complex than those that have them)
2. tense/aspect/modality forms (preverbal particles)
3. question words
4. a pluralizer
5. pronouns for all persons and numbers
6. a general locative preposition
7. an irrealis complementizer
8. a relativizing particle
9. reflexives and reciprocals

Grammatical functions that, if lost, will not be (immediately) reconstituted in the Creole:

1. gender agreement
2. number agreement
3. bound verbal morphology
4. derivational morphology
5. pronoun case and gender forms
6. almost all prepositions

Typical Creole Grammar (hedge: if emergence occurred catastrophically enough, with little retention of „un-creole“ material from superstrate or substrate)

- strict SVO word order
- case marking by position
- oblique cases marked with superstrate prepositions or serial verbs
- Wh movement is present, but NP movement is absent
- no participial or nominalized structures
- rare infinite structures
- semantics of grammatical words are highly constant (= sharply defined)
- functional words are recruited following well-known grammaticalization paths, and come from superstrate language (SPS)
- e.g. SPS for yields Creole irrealis complementizer, one yields indefinite article, demonstratives yield definite article, past form of copula yields past tense marker, finish yields a completive marker, some verb of location yields a completive marker, the verb
go yields irrealis/future, the question words are either copied or built systematically from SPS who or what plus words for place, time, person, etc.

Summary. In Bickerton, 1988 we face universal grammar in a „softer“ variant than Chomskyan UG. While principles and parameters are not discussed at any detail, Bickerton acknowledges universal pathways of grammaticalization as the canonical source for necessary grammatical words. Moreover, based on his survey of Creoles, he proposes a set of universally necessary grammatical functions that all natural languages need to cover. Escape hedges of the account: (a) type of creolisation; only rapid C. is described. (b) amount of retention which is, in part, driven by accident or unclear factors.

2.3. Measuring Complexity

When typological explorations of the languages of the world started on a larger scale, it was one of the big discoveries that scholars found no „primitive“ languages anywhere in the world. Without arguing in favour of a fine-graded scale of cultural advances, we have a good feeling that a culture of gatherers and hunters, with few and simple tools and techniques to exploit nature for food and shelter, is less advanced than a culture with agriculture, division of labour, extended trade connections, many tools and effective ways to make use of natural resources. The crucial findings of early typology were that

3. there were no similar differences in complexity between grammars of different cultures that would support the hypothesis that grammars undergo major qualitative developments and changes in historical time depth
4. more crucially, the complexity of grammar (and we may here think of complexity quite naively in terms of plausibility and perspicuity for the European scholar) did not in the least correspond to the degree of cultural advancedness of their speakers.

This fostered the belief that cultural development and linguistic development were distinct parameters and, for the sake of the argument, that all languages were in principle of equal complexity.

In spite of this general conviction, Creole languages were constantly compared to the grammar of the lexifier language. In the cases best documented, the Creole variety often was described as a degenerate variant of the superstrate, as is amply reflected in old names like Neger Hollands, Broken English or Nigger French. Only in the nineteenth century, scholars started to investigate Creoles in their own right, and a sharpened interest in the political and social implications of the classification of Creoles and Post-Creoles in the twentieth century led to more pronounced positions to the end that Creole languages are natural languages as valuable and well-designed as any other language (and specifically the superstrate). (Holm 2000 offers a balanced survey over the history of Creole studies.)

Against this historical background, a general scepticism against hypotheses to the end that Creole languages should still be qualitatively different from other languages and specifically be less complex comes as no surprise. One central point in the debate, obviously, is the notion of complexity and qualitative difference that is intended by different authors, at different times, and with different overall intentions. In the case of Creole languages, we may assert that not only is the grammar as mature as grammars can be, but also that speaker communities have no difficulty of expressing thoughts of any complexity, and exploit the same means as any other linguistic community to extend the lexicon by suitable items in order to match new and complex facts about the world that become important in everyday communication (e.g. metaphor, metonymy, sense extension, paraphrase, compounding, etc.). Since the ground-breaking work of Bickerton (1971) and as one of its implications, there is however a different notions of being less complex that is attributed to Creole grammars. The general perspective is that Creoles lack grammatical and phonological complexities that arise in the process of grammaticalization and phonological reductions in older languages. The positive
perspective of a language being less complex in this sense is perhaps best transported by paraphrases like streamlined, efficient, systematic, lacking marked constructions, or optimized.

If we assume the analysis of creole genesis along the lines of Bickerton—without necessarily subscribing to the Bioprogram Hypothesis in its stronger variants—it is a plausible expectation that newly-emerging Creole languages should lack several possible grammatical traits, always or at least in the majority of cases:

4. use of inflectional morphology
5. use of derivational morphology
6. differences in word order between main and subordinate clauses
7. marked phonemic systems (clicks, retroflexes, tones)
8. suppletive and other irregular paradigms of TAM-marking

The following research questions arise at this point:

- Does this expectation hold scrutiny if we look into a substantial number of Creole languages?
- Which are the actual grammatical traits that establish complexity? Is there a universal set of constructions, processes or features that show complexity, or should we think of a relative set of features?
- Are all Creoles non-complex in these dimensions?
- Are all non-Creoles complex in these dimensions? Is the lack of certain complexities only possible after Creolization? Is the degree to which Creoles (might) lack certain features only possible after Creolization?

McWhorter, 2001. Creoles are substantially less complex than „older languages“. McWhorter's first concern lies in invalidating the claim that all grammars are a priori of equal complexity. He points out that there is no pre-established mechanism that would ensure that losses in complexity at one end of a language would lead to more complexity at another end (unlike for instance the a priori law of Energieerhaltung in physics). Likewise there are no social, information-theoretical or other known laws so far that would suggest that competition between varieties tends to lead to the extinction of less complex variants, thereby guaranteeing an approximately equal level of complexity in all languages. The notion of complexity, first illustrated on an intuitive basis, is soon discussed and the author proposes the following dimensions that he will, tentatively, take into account.

A phonetic inventory is more complex to the extent that it has more marked members. Markedness is understood in the typological sense of being less frequent. Examples given are ejectives, clicks, labialized consonants as contrasted with stops, rounded back vowels, and glides. It rests on the Greenbergian notion of implicational relations, i.e. sounds count as marked if any language that possesses them also possesses all of the unmarked sounds. In particular, McWhorter assumes that a tone language is more complex than a non-tone language.

Comparing the syntax of two languages, that one counts as more complex that rests on more rules (e.g. asymmetries between matrix and subordinate clause; V2 in germanic languages). Ergative-absolutive languages are more complex than nominative/accusative languages; and a language that possesses more grammaticalized expressions for semantic/pragmatic distinctions counts as more complex than one with less such expressions.

Example: Koasati (Muskogean, Native America) with different existential verbs for five different types of object, depending on shape.

a. noo-k mat-hacá:l
   birch-SUBJ afar-stand
   „there is a birch over there“

b. ó:la-k tallá:k,
Finally, inflectional morphology renders a grammar more complex in comparison to a grammar without inflectional morphology. Why? In principle, inflections express the same contents that other languages might express by word order, prepositions, postpositions etc. Yet, as McWhorter argues, inflection can lead to morphophonological processes that make a grammar complex, e.g. unla ut as in German, or other sound shifts where stems are affected due to the presence of inflectional categories. Moreover, inflectional processes often are based on declension classes, grammatical gender or other semantically opaque classifications of the lexicon. Suppletion is likewise viewed as a possible side effect of inflectional morphology and a language with many suppletive paradigms is more complex. For all these reasons, McWhorter argues that inflection in itself is an exponent of complexity in a given language. (Clearly, this will stand in favour of his cause because all Creoles systematically appear to lack inflectional morphology. This can be predicted by observing that a purely content-word based Pidgin will exactly filter out such functional categories.)

On basis of these criteria, McWhorter compares the grammar of Tsez (Northeast Caucasian, Nakh-Dagestaninan language) and the grammar of Saramaccan CE (late 17th century). Summarizing his findings, he concludes that Tsez is more complex than SCE in at least the following dimensions:

- more complex phoneme system (42 segments, incl. uvulars, pharyngeals, pharyngealized uvulars; ejectives, labialized consonants; distinction between long and short a. contrasts with 25 phonemes, consisting of less marked sounds.
- Richer and deeper morphophonemic rules in Tsez than in Saramaccan: While the author refers to research papers which substantiate this comparison, we may acknowledge that phonological rules in Tsez are more in number, less evident, wider operant and with more sub-rules and exceptions than in SCE. (p. 142)
- Four noun classes in Tsez which are operant in agreement marking. Agreement marking interacts with focus marking: Focus can be signalled by agreement between verb and focussed argument (absolute). Tsez is also an ergative language. SCE does not have inflectional morphology, nor free noun classifiers.
- Nouns in Tsez can occur with stem alternation before certain inflectional suffixes. Nouns in SCE do not show stem alternation.
- Distinct case marking for experiencer verbs in Tsez: Experiencer is marked in lative case. SCE does not possess any extra marking for experiencer arguments. (Side remark: lative marking is not restricted to inflected languages; we know of prepositional experiencer marking as well. Hence, this is not an implication of SCE’s non-inflecting nature.)
- Tsez has a large number of derivational suffixes to derive nouns, verbs, adjectives: specifically denominal and deverbal abstract nouns, a marker designating residents of a place, two denominal adjectival markers with alienable/inalienable distinction, etc. SCE has two derivational affixes in total. Moreover it uses reduplication to derive attributive adjectives and resultatives from transitive verbs, and non-productively: nouns from verbs (ex. p. 140). Some of the derivative constructions are expressed in SCE by serial verbs (i.e. in another grammaticalized fashion), however most must be covered by analytic expressions. Hence, SCE is not less expressive than Tsez—it just expresses more contents by transparent semantic constructions.
- Word order in questions: Tsez shows different word order processes for Wh-words for adjuncts and arguments. Saramaccan has uniform Wh-fronting.
- Tsez has evidential past tense marking, Saramaccan does not. (Once again, other languages like the Tibeto-Burman Lahu show that analytic languages can in principle have evidential markers.)
Tsez has some suppletive plurals and some suppletive transitive-intransitive pairs. Saramaccan shows only one case of suppletion: before the verb gó 'to go', the imperfective marker tá is expressed as nan: 'mi tá wáka' I am walking contrasts with 'mi nán gó' I am going.

Generally, Tsez grammar forces the speaker to specify semantic distinctions more often, and at a finer level of granularity, than SCE. The opposite constellation rarely or never holds between these two languages.

Is this comparison just? Further complexities of Saramaccan, hitherto unmentioned:

7. Instances of phonological assimilation and epenthesis. (Which is however not dangerous; in fact, scholars like Holm take the rise of morphophonological processes as indication for creolization after the pidgin stage.)

8. Definiteness marking with a determiner system copied from English

9. Distinction between two alternant negations (á and ná)

10. Some lexical distinctions carried by tone (examples unclear whether accent or tone).

This is, McWhorter asserts, not counter to his claim because (a) creoles are not predicted to lose all complexity in creolization (unlike the BioProgramm Hypothesis would predict in its strong variant), and (b) a Creole that is 300 years old should expectably have developed results from grammaticalization. We just expect less of these than in „old“ languages.

McWhorter proceeds to claim that this comparison could likewise have been conducted with other languages that he holds as close to the Creole Prototype (Sranan, Ndjuka, Tok Pisin, Bislama, Solomon Islands Pidgin, Torres Strait "Broken", Aboriginal Pidgin English, Sao Tomense Creole P, Principe P, Anobones P, Angolar P, Negerhollands CD, Baba Malay, and Papia P are listed). The author expressly states at this point that other Creoles, particularly ones that were in continued contact with the lexifier language (e.g. Plantation French Creoles), over time also borrow derivational morphology from the lexifier and hence are not considered in the comparison.

As another language with a "fearsomely elaborated" grammar, he finally quotes Kabardian (Northwest Caucasian) and refers to pertinent quotations that Caucasian languages are in general counted as extraordinarily complex by any linguistic standard. While this undisputed observation has never raised critique and hence supports McWhorter's claim about the principled un-equality of languages in terms of complexity, one may doubt the justness of comparing Creoles with the attested peak of complexity in natural languages.

In the next section, McWhorter addresses the possible objection that analytical languages might be generally lower on his complexity scale than inflectional languages. If Creoles are analytical, then their low complexity would follow but not be a special feature of Creole grammars. In order to refute this hypothesis, McWhorter compares an analytic language, Lahu (Tibeto-Burman) and Saramaccan CE. Once again, the comparison sets Lahu higher on the scale in at least eight parameters: Phoneme System, Tones, Derivational morphology, Numeral classifiers, grammatical relations, modal and pragmatic particles (Abtönungspartikeln), number and use of serial verbs, verb compounds ("verb concatenation"). (pp. 146-148)

Such comparisons necessarily are isolated, and could offer an imbalanced score if the author was prejudiced. McWhorter, however, invites counter-scoring and predicts that no similar comparison could be devised that points into the opposite direction. More possible comparisons are invited in the same spirit, for instance a comparison SCE with English or, later in the paper and in a briefer manner than the two preceding comparisons, a scoring between SCE and Maori (analytic, and without tone).

At this point, the reader suspects that the fairest and most relevant comparison might be the one between substrate/superstrates and the emerged Creole. McWhorter indeed quotes the hypothesis of Michel DeGraff who claims that creole grammars should best be understood as resulting from their source language (super- or substrate??) in second language acquisition which typically shows

- loss of inflection morpholoy
subsidiary loss of syntactic processes and rules
subsidiary results due to the filtering-out of low frequency features
loss of certain functional distinctions

Importantly, DeGraff proposes that such losses and simplifications generally result from language contact, no matter whether Creole or other languages are involved. This sets his position clearly apart from that of the author. A full paragraph is devoted to clarifying that he thinks that all languages show simplification in language change, as well as acquiring complexity only Creole languages in the early years (200-300?) show a stage which is practically freed from all complexities.

One challenging case counter to this rigid claim is Riau Indonesian, a language that not only possesses one of the simplest phonemic inventories, but is also quoted to have practically no case marking whatsoever and hence is in some respects close to Pidgin level. (See below.)

At the end of the article, it is once again stressed that the validity of the main claim rests to a large degree on the notion of complexity. Several earlier versions of the hypotheses were doomed to fail, due to a poorly chosen notion of complexity. Specifically, the following spell-outs are refuted:

Less complex means semantically more transparent.
It may be tempting to assume that function words and items of an early stage still relate closely to their more contentful origin and therefore are used in a perspicuous manner. However, this advantage may be counterbalanced by the fact that small vocabularies also invite sense extension and generalization of words to cover more possible propositions (a process well-known in early L1 acquisition). Even though McWhorter acknowledges the underlying rationale as sound and valid, he stresses that his perspective does not rest on semantic transparency. (One might add that semantic transparency seems difficult to operationalize)

Less complex means less marked in terms of the Parameter Based Model of Grammar
It is explicitly refuted that the simplicity of Creole languages corresponds to unmarked parameter settings in UG (contra the early claims of Bickerton). In order to illustrate the independence of the two notion, McWhorter devises hypothetical grammars that are 'simple' in his terms but 'complex' in terms of parameters in X'-syntax, and vice versa grammars that are 'complex' in terms of parameter settings but 'simple' in his metric. Essentially, this is due to the fact that many aspects of McWhorter's metric concern the morphophonological side of grammar, phonology, or parts of the lexicon (of functional items like particles) which are not covered in the Principles-and-Parameters framework.

Less complex means lower on an implicational hierarchy in the Greenbergian sense.
This is a perspective which is explicitly embraced by the author. However, his method of a free comparison of grammars may allow for finer comparisons between languages than a universal catalogue of grammatical traits.

While McWhorter attempts to operationalize the claim of his paper, the essential underlying idea—or, if we wish, the positive conviction that such a claim might come out true—consists in the following basic picture:

Grammaticalization, in many cases, leads from one non-defective grammar to another non-defective grammar with one "rule" or "construction" or "morpheme" or "process" etc. in addition to those that were present at the earlier stage.

While from time to time, this accumulation of grammar leads to simplifications by a complete reorganization of grammar (like e.g. the emergence of modals, word order change in English, etc.) and sporadically, exceptions are reduced by analogy to more frequent and regular patterns, McWhorter's perspective is grounded on the impression that under conservative language
development, these reorganizations and simplifications don't counterbalance the opposite processes.
Interestingly, if we trust the democratic vote of the community of scholars in grammaticalization and language diachrony, notions like "Language Change as Optimalization" are hotly debated and authors like Lightfoot, Roberts, van Gelderen etc. under constant attack for such views. In those quarters, it appears a completely open question whether results of grammaticalization make a language—even locally—a better, more balanced system or whether categories like improvement (or in the other direction: unclear complication) should not be used at all.


Varieties of Riau:
Non-mainstream isolects — Mainstream isolects — Indonesian isolects (branching out to Standard Indonesian) — Outsider isolects.
Riau Indonesian: Indonesian isolect; used for inter-group communication, but a vernacular rather than the official standard language.
Emerged in century-long migration e.g. of the Minankabau (Malay varieties) from West Sumatra, through the Riau province (more Malay varieties), on to Malaysia.
Further contact influence by migrants of Chinese origin.
Varieties spoken within Riau province reflect all the grammatical traits that can be found looking at the full set of varieties of Malay languages.

=> Koinezation. Permanent mixing, but no indications or legends of colonialization, Creole situation. (Gil is careful to argue against any version of possible Creole origin of Riau Indonesian. He can rely on an independent characterization of „Pidgin Malay Derived“ languages by Adelaar and Prentice, and shows that their characteristics do not fit RI.)

Gil proceeds to a comparison between RI and Saramaccan, using the catalogue of features defined by McWhorter:
Phonology: A language is more complex than another to the extent that its phonology has more marked members
Syntax: A language is more complex than another to the extent that its syntax requires the processing of more rules
Grammaticalization: A language is more complex than another to the extent that it gives overt and grammaticalized expression to more fine-grained semantic or pragmatic distinctions
Morphology: A language is more complex than another to the extent that it has inflectional morphology.

Briefly, Gil argues on basis of these criteria that Riau Indonesian is less complex than the Creole Saramaccan, and considerably so. (For details see original article; e.g. table 4 „syntax“ on page 344.)

He proposes that all Creoles may be simple, but that not everything which is simple needs to be a Creole. Specifically, he proposes that RI has profited (if you like languages of little complexity)
- from constant migration and L2 acquisition,
- from its basilectal status
- from contact situation of a family of isolating languages (Chinese, Vietnamese) with phonologically simple languages (Malay family).

**2.4. Complexity: Staying within your family**
Problem of universal claims like Bickerton, McWhorter (2001): Comparison of complexity between languages of completely different language families.

Peter Trudgill


<table>
<thead>
<tr>
<th>Waldeck dialect</th>
<th>earlier form</th>
<th>Standard German</th>
</tr>
</thead>
<tbody>
<tr>
<td>iks</td>
<td>*i:s</td>
<td>Eis</td>
</tr>
<tr>
<td>driksiç</td>
<td>*dri:siç</td>
<td>dreißig</td>
</tr>
<tr>
<td>bìksen</td>
<td>*bi:sen</td>
<td>beißen</td>
</tr>
<tr>
<td>pikfe</td>
<td>*pi:fe</td>
<td>Pfeife</td>
</tr>
<tr>
<td>likp</td>
<td>*li:p</td>
<td>Leib</td>
</tr>
<tr>
<td>nige</td>
<td>*ni:ne</td>
<td>neun</td>
</tr>
<tr>
<td>fukst</td>
<td>*fu:st</td>
<td>Faust</td>
</tr>
<tr>
<td>uks</td>
<td>*u:s</td>
<td>aus</td>
</tr>
</tbody>
</table>

... 

Metatypy: Contact phenomenon in open but tightknit polylectal communities: Speakers of a language A retain the original, language specific lexicon while borrowing the syntax / grammar of some influential contact language B. (In terms of creolization, one would expect that B was a substrate of a Creole based on lexifier A.) Ross quotes examples where Creolization, according to all we can say today, has not played a role in the mixing of two languages. Example

Karkar island, north of Papua New Guinea. Two communities: North: Waskia—Papuan lexicon, but grammar/syntax looks like Austronesian languages. South: Takia—clearly an Austronesian language on all accounts. Ross: Waskia became reformed in constant language contact. There was no creolization of an Austronesian community by a superstrate Papuan society, as far as the records tell.

John McWhorter, 2005:
Comparison of languages within one language family. Textbook example: English in the Germanic languages.

Claim 1: Language families can have a „black sheep“, one member that is systematically simpler in all respects than its sibling languages.

Table of exceptional simplifications in English, compared to other Germanic languages
be-perfect | passive become | V2
---|---|---


More examples: Persian in the Indo-Iranian family, Mandarin Chinese in the sinitic language family.

**Claim 2:** These simplest members come about iff there is massive immigration or contact by speakers of a related variety who perform imperfect second language acquisition, acquiring a variety that is the „smallest common denominator“ of the two related varieties. Nonliteral society. (See also the Riau Indonesian situation!). Example for English: Invasion by Scandinavians, influence by second language acquisition, koineization, Old Norse influence.

**So far** these are claims that (once again) elicit controversial discussion.

**Challenging corollary:** Looking at the time scale of the English case, one would have to conclude that massive L2 acquisition can shred down a language to a degree where it does not even recover by grammaticalization in a stretch of 1000 years...

### 2.5. Overall Summary

Details are still unclear, but the following trends are attested and accepted.

1. Substantial parts of creole languages emerge „out of nothing“ in L1 acquisition. It is possible that all grammar starts out of nothing. (Unlike other cultural achievements, NL will never ever get lost. As long as we reproduce and raise children, we will also speak.)
2. The result is tendentially structurally simpler than „old“ natural languages
3. Yet, the following extreme positions are viewed with scepticism:
   (a) „Creoles are substantially more simple than any other languages“.
   (b) „Creoles witness Universal Grammar in the sense of __________“ (insert prominent author of your choice).
4. The following further factors that support language „simplification“ have been proposed: (a) High rates of L2 acquisition; possibly or preferably by speakers of a related L1 variety. (b) Koinezation: High rates of L2 acquisition by other migrants over prolonged periods (Riau Indonesian).
5. Complexity arises in isolated language development. Socio-Linguistic factors may support the maintenance and stressing of special traits of small, tightly-knit language communities.
Part 3: Language Change as Evolutionary Process
Formal and Computational Models

There is a longstanding tradition of research into the analogy between the evolution of species and the evolution of languages. For X set *language* or *species*:

- Variation within one X.
- Variation can be strengthened between different, separated populations of (speakers of) X.
- Long and divergent variation can lead X to develop into two distinct descendand X’es.
  *species*: no interbreeding. *languages*: no mutual comprehensibility
- Diachronic development of family of X’s mirrored in family trees.

Does the parallel extend to the existence of selective features?

X = *species*

In new environments, fitter variants can have a reproduction benefit. There are external forces that drive the population towards new genetic „shapes“.

X = *language*: Grammars (I-language) reproduce through language acquisition. Primary Linguistic Data (PLD) produced by adult speakers provide input for learner generation and „create“ new I-grammars.

Assuming that we start with a generation 0 of homogeneous speakers
- How do the first variants in I-grammar come about (genetic variation?)
- Do certain I-grammars have reproductive advantages, under certain circumstances, over other, competing I-grammars?
- What are the environmental factors that influence selection of certain I-grammars over others?
- Is there any sense of talking about *fittest* grammars?

*Biology*: We have reached a good understanding of the interconnection between genotype, phenotype, and the kind of environmental factors that might be relevant (successive competition for food, shelter, mating, … => reproductive success).

*Language*: Less clear landscape of environmental factors and reproductive success. Good grammars are?
- structurally simpler than their competitors?
- more prestigious than their competitors?
- more valued within one’s ethnic group of reference (although perhaps not overall prestigious)?
- better to convey information to as many speakers as possible?
- better to hide information from as many bystanders as possible?
- acquired on the basis of more robust cues?
- … ?

**3.1. A new liaison between biology and linguistics**

Computational models of reproductive systems: Simulate the long term effects of certain reproductive factors in a given system, using sophisticated models of learners embedded in a dynamix context. (Verbal theorizing about the likely behaviour of complex dynamical systems is often mistaken, as we don’t seem to possess good intuitions about such dynamic developments.)
Typically, models consist of the following components:

1. A population of agents (e.g. child acquirers turning into adult speakers)
2. A space of possible signals (usually strings of symbols).
3. A space of possible meanings (usually some kind of structured representation in some quasi-logical notation; such that some meanings are more similar to each other than others).
4. A production model. This determines how, when prompted with a meaning, an agent uses its knowledge of language to produce a signal.
5. A learning model. The learning model defines how an individual agent acquires its knowledge of language from observing meaning-signal pairs produced by other agents.


More detailed variants of this overall setup will be encountered in the sequel.

SWARM software: www.swarm.org/wiki/Introduction_to_swarm

Note: In this part of the lecture, the word „evolution“ will be encountered frequently. Apart from some isolated occasions, this will NOT refer to the evolution of language within the species homo sapiens. The prime interest will NOT be what induced the first hominides to talk.

The new liaison between biology and linguistics takes issue with languages as self-reproducing systems and investigates the factors of language development, GIVEN that the language faculty, as well as fully functioning natural languages, are already existent!


3.2. Learning the Word.


3.3. Why Languages Better had a Grammar.


Authors build on the Chomskyan paradigm of language. They stress that the difference between „learning“ and „memorization“ is the ability to generalize beyond one’s own experience to novel circumstances. Language acquisition poses the challenge of the poverty of stimulus: On the basis of limited input, the child is supposed to choose the right one within a range of possible target grammars. Universal expectations about the intended grammar (UG) are a formal necessity:
There exists no algorithm that can learn the set of regular languages (Gold, 1967). Per implication, no algorithm can learn the set of all context free, or context sensitive, or computable languages. Critics have proposed more specific, or different, spaces of possible grammars to be searched (statistical learning; cognitive grammar). All these, however, can be seen as one way to limit the search space by the principles of UG.

UG defines a set of finitely or infinitely possible grammars \( \{ G_1, G_2, G_3, \ldots \} \).

Nowak’s research question is one about the evolution of the language capacity as such (!): What makes a population of speakers converge to a coherent grammatical system? What are the conditions that UG has to fulfil for a population of individuals to evolve coherent communication?

Authors envisage a heterogeneous population of speakers who use slightly different languages. Those who are better at the language game are assumed to have a reproductive advantage. (Note, once again, that Komarova and Nowak investigate the biological evolution of the UG capacity, and not language internal evolution.)

The model that the authors devise links the following factors:

- the range of possible grammars \( G_i \) (all within the limit of the UG) that are spread in a given population
- the probability \( s_{ij} \) that a speaker who uses grammar \( G_i \) says something that can be understood by a speaker of grammar \( G_j \)
- the mutual understandability between two grammars \( F(G_i, G_j) \) which is computed on the basis of the above
- the proportion \( x_i \) of speakers of \( G_i \) in the community
- the average pay-off for such an individual to speak \( G_i \) (i.e. the probability that it will, in average, be understood by, and understand, speakers of other grammars)
- a stochastic matrix that reflects the learnability of grammars by children: \( q_{ij} \) lists the probability that a child of parents with grammar \( G_i \) ends as a speaker of grammar \( G_j \). Note that in the described minimal scenario, the child only gets input from one parent. The model hence measures the mere learnability of a grammar on the basis of coherent but limited input. It does not treat competing input from several grammars; although the authors assert that one could.

Which properties must UG have such that a predominating grammar will evolve in a population of speakers? Which UG can induce grammatical coherence in a population?

Two possible learning scenarios are explored:

a. Memoryless learning algorithm: Learner uses grammar \( G_i \) until it meets conflicting evidence. It will then pick another grammar \( G_j \) at random. Changes occur until the chosen grammar conforms with the input.

b. Batch learner: Memorizes all sentences and at the end chooses the candidate grammar that is most compatible with the input.

First, simplifying assumption: All languages are equidistant from one another. (No language is more likely to be understood by other speakers than any other): \( s_{ij} = s \) constant for \( i \neq j \); and \( s_{ii} = 1 \).

Memoryless learner: Grammatical coherence can emerge if the number of input sentences in acquisition, \( N \), exceeds a constant times the number of candidate grammars: \( N > c_1 \cdot n \).

Batch learner: Grammatical coherence emerges if the number of input sentence \( N \) exceeds \( N > c_2 \cdot \log n \).
More general case: Distance between languages is given by a uniform distribution.

Memoryless learner: Grammatical coherence can emerge if the number of input sentences in acquisition, \( N \), exceeds \( N > c_1 \cdot n \cdot \log n \).

Batch learner: Grammatical coherence emerges if the number of input sentence \( N \) exceeds \( N > c_2 \cdot n \).

\[ \Rightarrow \text{Coherence threshold: Limits the size of the search space relative to the amount of input available to the child. (As long as no sufficiently restrictive UG has evolved in the brain of these children, the population as a whole will not be able to cohere on some grammar.)} \]

Humans will perform better than the memoryless learner, but worse than the batch learner:

\[ c_1 \cdot n \cdot \log n > \text{human threshold} > c_2 \cdot n \]

In prose: UG has to evolve to a stage where the number of possible grammars \( n \) is small enough to ensure individual learning and common coherence in the community within the limits set by the lifetime parameters of individuals.

**List grammars and rule based grammars**

Under what circumstances should nature favour rule based grammars over list based grammars? The authors consider a group of individuals that use \( M \) different sentences (syntactic structures). \( M \) specifies the number of sentences that are relevant from the perspective of biological fitness.

\textit{List learners} need to hear at minimum \( N \geq M \) sentences in order to acquire the language. 

\textit{Batch learners of a grammatical rule system} (with grammar similarity coefficients \( s_{ij} \) are distributed uniformly between 0 and 1) need to hear at minimum \( M > c_3 \cdot n \) sentences in order to acquire their grammar.

In prose: If there are many more important messages \( M \) than choices of one target grammar from \( n \) possibilities in UG, then the evolution of such a restrictive UG is selected for.

**Summary and Criticisms:**

The work by Nowak and colleagues explores the reproductive influence of subsequently evolving systems of UG, making use of dynamic replication equations known from biological reproduction. The authors derive several threshold values for innate UG systems from when on such an innate trait will become evolutionary advantageous. We envisage generations of hominides with some proto-languages where the emergence of the potential to acquire and deliver a coherent common grammar from a certain point on acts as a reproductive advantage.

Two aspects of doubt should be expressed here. Firstly, it is a common experience that the lack of a common lexicon impedes communication to an infinitely larger degree than the lack of a coherent consistent grammar. It is for good reason that \textit{linguae francae} possess a fixed lexicon with little or no grammatical structure, and not vice versa. The modelling of the authors fails to comment on the question why the sign-meaning association is not restricted genetically in the least (i.e. requires tedious learning even in those aspects which we can assume to be common to all human experience) while grammatical coherence is presented as the fatal reproductive factor.

Secondly, filling the abstract considerations about thresholds with known numbers might shed a new light on the results offered. The threshold between a \textit{list learner strategy} and a \textit{batch grammar learner strategy} states that batch learners do better if it is advantageous to convey many more important message types (syntactic structures, \textit{not} lexical content!) than the range set by possible grammars in UG. Taking into account that UG must comprise at least the grammars of all
living languages (app. 6000) and presumably more (such that even 10,000 may be a reasonable finite count for the range of possible grammars selected by UG), we get as a corollary that

*It was of reproductive advantage for speakers to be able to convey (much?) more than at least 10,000 distinct grammatical structures. This created the selective pressure from list learners towards rule learners.*

Without wanting to go into these figures, such a scenario is doubtful. It might be an illustrative exercise to explore the range of grammatically distinct constructions in e.g. Riau Indonesian in order to get a feeling for the plausibility of such numbers.

Nowak et al. might argue that the reproductive advantage consists in improved mating chances (due to impressive eloquency) rather than improved survival rates in terms of finding shelter and food (where less syntactic constructions are plausibly sufficient). However, as Tecumseh Fitch points out (T. Fitch, 2005: A Biological Perspective on the Evolution of Language. Talk presented at Blankensee Colloquium 2005, Berlin), the biological facts of language acquisition point straightforwardly against the view of language as a mating signal. Specifically, Mating signals always get prominent when individuals reach maturity—the language learning capacity closes down exactly at that time Mating signals are usually distinct for the two sexes—the language capacity is equally developed for either sex.

The following approach to the genesis (or, advantageousness) of compositionality, although not designed to span the development from early hominides to homo sapiens, rests on safer ground.

**Simon Kirby. 2002. Learning, bottlenecks and the evolution of recursive syntax.**

Language in two different domains: I-language and E-language. (*langue* and *parole*, *grammar* and *speech*, …).

I-language of generation n → E-language produced by generation n

= Input for acquisition of generation n+1

→ I-language of generation n+1

Simulation implements the following processes:

1. An individual in the simulation is given a set of meanings that must be expressed. These meanings can be thought of as being procided by the external 'world', but in the simulation are simply chosen randomly from some predefined set.
2. The individual then attempts to express each meaning either using their own internalized knowledge of language or by some random invention of a new 'word' for the new meaning.
3. A new learner takes such sets of utterances and uses them as input to learning.
4. Finally, the learner becomes a new speaker, the old speaker is discarded ('dies') and a new individual is added to become a new learner, and the cycle repeats.

Representation of form-meaning pairs: 'words' and 'phrases' are strings of letters (simplifying vastly about phonology). 'meanings' come in a quasi predicate-logical format. Example for something like an English utterance:

<

tigereatsjohn ;

EATS(TIGER,JOHN)
>

The learning algorithm:

**Possible grammars.**

Search space for possible adult grammars is assumed to consist of context-free grammars, with semantic content attached to syntactic units (of any complexity). Complex meanings can be
attributed holistically to full utterances. The grammatical search space will, however, also provide grammars that use compositionality. (See example grammars on page 178).

**Rule subsumption.**
Firstly, the learner holds no rules. The generation 0 speaker will pair meanings with strings according to arbitrary invention. (‘Adam stage’). E.g.:

```
<tigereatssausages ; EATS( TIGER, SAUSAGES)>
```
as

```
S/ EATS( TIGER, SAUSAGES) \rightarrow tigereatssausages
```

**Trivial learning algorithm:** The dictionary method. Collect all possible utterance-meaning pairs without further analysis.

**More sophisticated learning algorithm:** Search for generalization patterns. „What is the least general rule that subsumes a set of given rules?“: E.g.

```
S/ EATS( TIGER, SAUSAGES) \rightarrow tigereatssausages
S/ EATS( JOHN, SAUSAGES) \rightarrow johneatssausages
S/EATS(X, SAUSAGES) \rightarrow N/X eatssausages
N/TIGER \rightarrow tiger
N/JOHN \rightarrow john
```

**More subsumption rules:** Minimizing grammatical categories. E.g.

```
N/MARY \rightarrow mary
M/MARY \rightarrow mary
```

will be simplified by choosing one of the non-terminal category symbols M or N, and replace all occurrences of the non-chosen symbol in the grammar by the chosen one.

**Induction algorithm**
Take an utterance. Incorporate the simplest possible rule that generates this utterance directly. Search existing rule space for possible subsumptions like those described above, until no further generalizations can be found. Delete duplicate rules. Proceed to next utterance.

**Invention:** What can a speaker do in order to express a new meaning?
Take meaning that is to be expressed. Try to find the closest meaning that the speaker *does* have a way of producing. Generate parse tree, and spot mismatching parts: Pieces of meaning that do not match the intended meaning *plus* the parts in the parse tree where surface form contributes that meaning. Replace that part by a random expression, assigned to the intended meaning of that subpart.

If no such invention of new words for parts of a meaning can produce the new meaning within the range of the existing grammar, it is also possible to create a new word for the intended meaning holistically.

**Summary of the simulation cycle:**

1. The speaker tries to produce a set of utterances that will form input to the learner. This involves the following sequence of actions, repeated some number of times:
   - The speaker is given a meaning chosen at random from a predefined set.
   - If the speaker is able to generate a string for that meaning using its grammar, it does so, otherwise it invents a string. If the speaker has invented a string, the *speaker* uses that string-meaning pair as input to induction. This means that, if an individual invents a new way of saying something, they will learn from that and use that invention if necessary.
The learner is given the string, and tries to parse it with any grammar it may have. If it is unable to parse the string, then it takes the string-meaning pair and uses it as input to induction.

2. The speaker’s grammar is logged and then it is deleted from the simulation.

3. The learner becomes the new speaker, and a new learner with a blank grammar is added to the simulation.

Points of experimental variation: Number of utterances that speaker can produce during its lifetime, and structure and size of meaning space.

A. Experimental results: Development of simple compositional grammars.
Simulation starts from a simple finite meaning space of binary predicates and individual concepts. No reflexives were allowed.
Each speaker expresses at most half of the possible meanings in the meaning space during its lifetime (‘bottleneck’), randomly chosen.
In the early stages, all systems have low expressivity (= don’t exhaust the full meaning space) and large numbers of rules.
Languages universally moved towards higher expressivity and smaller grammars.
In the end, systems develop highly regular grammars reflecting distinct categories for predicates (‘verb’) and entities (‘nouns’). Passing through stages of apparent free word order, typically stabilizing on one fixed word order eventually.
(Simulation example see pp. 185-188)

B. Experimental results: Infinite meaning spaces
Simulation allows embedded meanings, thereby creating an infinite meaning space.
Agents start by producing 50 form-meaning pairs to name individual concepts (‘degree-0 meaning’), 50 form-meaning pairs for simple sentences (‘degree-1 meaning’) as well as 50 for meanings that involve recursion (‘degree-2 meaning’).
In the early stages, grammars have large numbers of rules (up to 200) and E-language typically contains a substantial proportion of invented form-meaning pairs.
Again, grammars reduce the number of rules to about 15-20 (after app. 100-200 generations).
Compositional grammars emerge that contain rules for recursive embeddings (see ex. p. 191).

Discussion: The fact that learners have only access to limited experience from a potentially larger or unlimited space of meanings acts as a ‘bottleneck’ for language transmission. Only grammars that use compositional strategies will ‘Survive’ eventually. Even if there were individual speakers with large non-compositional grammars (which can effectively map the meaning space into language) these grammars have no chance to be transmitted to the next generation.
Compositional grammars emerged even though no external pressure was implemented (to the end that only those speakers will survive who are in command of the larger meaning space).

Note that the implemented acquisition algorithm, although not defining a specific grammar or even set of grammatical categories, is not a blank sheet. Compositionality as a strategy is predefined in the algorithm. Yet, it is elucidating that the mere existence of this strategy (e.g. as the ability to decompose and generalize) is sufficient for a compositional grammatical system to develop over time.


• Strikingly naturalistic simulations of phonological erosion in interaction with frequency of utterance: Utterances of adult generation are not randomly chosen but following a certain non-equal probability distribution. Pronological erosion is simulated by random loss of single letters in lexical items, in isolated utterances. Result: After several generations, there is a clear match between the frequency of of a word in use, and its length: The more frequent a word, the shorter it is.
• Striking interaction of putative independent factors (p.c.): In one series of simulations, utterances were afforded with a toy CV syllable structure. Surprisingly, syllable structure was interacting with preferred word orders of sentences. While other setups yielded SOV grammars, SVO grammars OVS grammars, OSV grammars etc. with random distribution (and even mixed grammars), syllable structure set a trend towards a reliable N V N sentence type.

3.4. Struggling with the Facts.


Niyogi offers an application of dynamic models of cultural transmission to the case of language acquisition. Specifically, the change of English word order (see part 1) is treated as a sample study.

3.4.1. The setup of general models of cultural transmission

The Cavalli-Sforza and Feldman (1981) model starts from parental generation with a culturally transmitted feature in two values, e.g. C (carnivore) and V (vegetarian). Each individual is assumed to have exactly one of these two values, the values are transmitted by learning. Children acquire their value from the parental environment.

<table>
<thead>
<tr>
<th>Father</th>
<th>Mother</th>
<th>p(Child = V)</th>
<th>p(Types)</th>
<th>random mating</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>C</td>
<td>$b_3$</td>
<td>$p_3$</td>
<td>$(u_t)^2$</td>
</tr>
<tr>
<td>C</td>
<td>V</td>
<td>$b_2$</td>
<td>$p_2$</td>
<td>$u_t(1-u_t)$</td>
</tr>
<tr>
<td>V</td>
<td>C</td>
<td>$b_1$</td>
<td>$p_1$</td>
<td>$u_t(1-u_t)$</td>
</tr>
<tr>
<td>V</td>
<td>V</td>
<td>$b_0$</td>
<td>$p_0$</td>
<td>$(1-u_t)^2$</td>
</tr>
</tbody>
</table>

$C =$ carnivore, $V =$ vegetarian

$h_i =$ probability for child of given parental type to become vegetarian

$p_i =$ probability of parental type in population (at generation 1)

$t =$ generation number

$u_t =$ proportion of people being vegetarians (V) in generation $t$

random mating: probability of parental type in next generation

Proportion of V types (vegetarians) in the population will develop according to the update rule

$$u_{t+1} = B(u_t)^2 + Cu_t + D$$

with $B = b_3 + b_0 - b_1 - b_2$; $C = b_2 + b_1 - 2b_0$, $D = b_0$

3.4.2. Application to language acquisition

Assuming that the parents speak a given language in two varieties $L1$ and $L2$, the transmission probability for children’s grammars depend on the learning algorithm employed by children.

Let $L1, L2$ be the two language varieties at issue.
Let $g_1, g_2$ be the two grammars that generate them.
Assume that speakers of $L1$ produce sentences of $L1$ according to a probability distribution $P1$, and speakers of $L2$ produce sentences of $L2$ with probability distribution $P2$. 
Let $D_k$ be the set of all subsets of utterances over the common lexicon of $L1, L2$ with length at most $k$. Each of these sets could be a candidate dataset that could serve as the (finite) primary linguistic data for a child in language acquisition. A learning algorithm $A$ is a computable mapping from $\cup_k D_k$ to \{g1, g2\}.

Niyogi initially refers to the Cavalli-Sforza and Feldman algorithm which implements the following learning scenario:

1. Children of parents who speak the same language receive examples only from the unique language their parents share (i.e. children of $L1$ parents only get $L1$ input, and likewise for $L2$).
2. Children of parents who speak different languages receive examples from an equal mixture of both languages, i.e. they receive examples drawn according to $1/2P1 + 1/2P2$.
3. After $k$ examples, language acquisition counts as finished. Children retain the grammar they have at that moment for the rest of their lives.

However, Niyogi and Berwick (1995, 1997, 1998) devise an alternative evolutionary model which is both simpler as well as less confined to a core family acquisition scenario. While the models sometimes differ in the speed of developments, both are shown to offer equal results for the interesting parameter settings. The Niyogi-Berwick scenario is the following:

1. The population can be divided into children and adults.
2. All children in the population are exposed to sentences drawn from the same distribution.
3. The distribution with which sentences are drawn depends upon the distribution of language speakers in the adult population.

For each generation $t$ of adult speakers, let $s_t$ be the proportion of $L1$ speakers in generation $t$.

On basis of these assumptions, we get the development of the proportion of $L1$ speakers in the population at generation, in dependence of the proportion in generation $t$, the learning algorithm $A$, the distributions $P1$ and $P2$, and the maturation time $k$:

$$s_{t+1} = f(s_t) = g(A, s_tP1 + (1-\ s_t)P2, k)$$

where $g(A,P,k)$ equals the overall probability that a learner, using algorithm $A$, will acquire grammar $g1$ (i.e. speak $L1$) when exposed to primary linguistic data of at most size $k$ assuming that the language is used according to a given probability distribution $P$. (equation see p. 210)

The probabilities of language usage will simplifying be assumed to be uniform distributions; in order to do any better, empirical observations in populations would be needed.

The functioning of the learning algorithm, in contrast, can be subject to theoretical scrutiny. The author turns to that issue in paragraph 3 of the paper.

3.4.3. The Triggering Learning Algorithm (Gibson and Wexler, 1994)

1. Initialize: Start with randomly chosen input grammar.
2. Receive next input sentence $s$.
3. If $s$ can be parsed under current hypothesis grammar, go to 2.
4. If $s$ cannot be parsed under current hypothesis grammar, choose another grammar uniformly at random
5. If $s$ can be parsed by new grammar, retain new grammar, else go back to old grammar.
6. Go to 2.

(Another of the memoryless learning algorithm, although this one allows for a little memory.)
An analysis of this learning algorithm as a Markov Chain allows to derive an actual computation of the function $f(s_t)$ and hence compute the proportion of $L1$ speakers in generation $t+1$ from that in generation $t$. (For actual equations, see original article.)

The evolutionary system turns out to hinge crucially, and only, on the following two parameters:

- $a =$ the probability with which a speaker of $L1$ produces sentences $s$ that might be $L1$ or $L2$ sentences (ambiguous sentences).
- $b =$ the probability with which a speaker of $L2$ produces sentences $s$ that might be $L1$ or $L2$ sentences (ambiguous sentences).

### 3.4.4. Long term development

The model allows variation at the following points: How many sentences does a child see before maturation? How many ambiguous sentences are produced by speakers of languages $L1$, and $L2$ respectively? (Note that the probability distributions in all other respects cancel out in the final equations.) For all parameter settings, we want to know how the system will develop after arbitrarily many generations of speakers, i.e. in the long run. Will it be an $L1$ community, or an $L2$ community, or a mixed community, and if so, at which proportions?

Niyogi discusses the interesting limit case of $k \to \text{infinitely many sentences of acquisition input.}$ (Under this scenario, it turns out that several competing models that are issued in the paper show similar behaviour.)

For $a=b$, the child generation has exactly the same proportion of $L1$ speakers as the parent generation, i.e. there is no change.

For $a>b$, the population composition $s_t$ tends to 0. In prose: If speakers of $L1$ are more likely to utter ambiguous sentences than speakers of $L2$, then over time, the community will turn into an $L2$ community.

For $a<b$, the population composition $s_t$ tends to 1. In prose: If speakers of $L2$ are more likely to utter ambiguous sentences than speakers of $L1$, then over time, the community will turn into an $L1$ community.

(Refinements of the system with bi-lingual speakers that use both grammars in mixed proportion are possible. The behaviour of the system for arbitrary $k$ could not so far be characterized mathematically. The behaviour for $k=2$ is known.)

### 3.4.5. The change of English Word Order.

Niyogi moves on to an actual case study, the replacement of English SOV by SVO word order (for data, see Part 1). For ease of discussion the OV/VO parameter and the I-final / I-medial parameter are collapsed into a single head-complement direction parameter. Additionally, the V2 parameter (V2 movement or not) is taken into account. Niyogi refers to Lightfoot’s hypothesis according to which a certain threshold proportion of triggering data (about 30 %, judging from contemporary Dutch V2 PLD) are necessary for a V2 SOV grammar to be acquired.

#### Case study 1: ±V2 for head-initial grammars.

Initial degree-0 sentences (without recursion) of $L1$: no V2 movement, head-first, Spec-first:

$L1 = \{ SV, SVO, SV\ O1\ O2, S\ Aux\ V, S\ Aux\ V\ O, S\ Aux, V\ O1, O2, Adv\ S\ V, Adv\ S\ V\ O, Adv\ S\ V\ O1, O2, Adv\ S\ Aux, V, Adv\ S\ Aux\ V, Adv\ S\ Aux\ V\ O, Adv\ S\ V\ O1, O2\}$

(corresponds to modern English; e.g. $S\ Aux\ V\ O1\ O2$ maps to $John\ will\ eat\ beef\ in\ London.$)

Initial degree-0 sentences (without recursion) of $L2$: V2 movement, head-first, Spec-first:

(Obligatory movement of inflected verb to second position; one error Adv S V corrected to Adv V S here).

Ambiguous patterns: $$L_1 \cap L_2 = \{ \text{S V, S V O, S V O1 O2, S Aux V, S Aux V O, A Aux V O1 O2} \}$$

Additional assumption: All sentences are used with uniform likelihood by all speakers.

Result:

$$a = \frac{1}{2}$$ (ratio of ambiguous sentences in $$L_1$$)
$$b = \frac{1}{3}$$ (ratio of ambiguous sentences in $$L_2$$)

Prediction: $$L_1$$ (the non-V2 variant) should die out.

This stands counter to the actual development in the history of English.

Case study II: OV or VO for +V2 grammars.

Once again, the degree-0 sentence patterns of $$L_1$$ ($$V_2, \text{Head-first, Spec-first}$$) and $$L_2$$ ($$V_2, \text{Head-final, Spec-first}$$) are listed. (p. 224)

From that, Niyogi can compute the intersection of patterns, and the ratio of ambiguous sentence patterns in $$L_1$$ and $$L_2$$. As a result, we get that $$a = b = \frac{1}{2}$$. That is, the resulting system should show no particular trend towards either language type.

Both case studies do not offer a particularly strong analysis into the historical development of English. Niyogi stresses that they are meant as illustration for the underlying algorithms and equations rather than as fully-fledged analyses. (Section about geographic neighborhood effects is skipped here.)

Discussion:

• Problem of initial variation: The account can not explain how the first variation should arise in a homogeneous $$L_1$$ community. Even if parts of $$L_1$$ input could be $$L_2$$ input as well, learners will never see any unambiguous $$L_2$$ but occasionally unambiguous $$L_1$$ input. This should shift the equations so as to always ensure $$L_1$$ to $$L_1$$ acquisition for large $$k$$.

• Predictions of account could be improved if taking care of discourse value of different sentence patterns. Unambiguous $$L_2$$ sentences may recede in use if speakers fail to see, or cease to use the special discourse functions that were carried.

• Modelling is however crudely inadequate in present form, due to the crucial status of embedded sentences for acquisition of Head-final variety. Likewise, split of Aux - V ("verb bracket") remains undiscussed.

• Further explorations of the model necessary in order to prove its usefulness.

3.5. Can Principles be Acquired?


Is language / Is Universal Grammar and all-or-nothing ability? How can it ever have been selected for?

General evolutionary landscapes:
a. The gradual model. If an ability A can build up from zero over intermediary stages, and if each intermediary stage makes its holders „fitter“ than the competitors, then evolutionary theory offers a straight road towards ability A. Pinker and Bloom (1990) argue that language capacity must be, for that reason, a gradually accumulated ability. Otherwise we are in option b. The all-or-nothing model. If an ability A is complex, but is present all, or not, then there is no selective force which would drive genetic development towards this isolated adaptive peak. There is no genetic trend to accumulate adaptions towards that peak. For language, one then would assume that it is rather the by-product of some other genetic adaption (a preadaptation or exaptation).

Learning guided evolution:
As an intermediary possibility, the Baldwin Effect (after a classical paper by J. Mark Baldwin, 1896) has been discussed. It assumes a two-stage development. Organisms may be able to show learned behaviour in addition to structures fixed by the genome. In changing environment, those with more ability to learn how to cope with the new environment will survive while those without are selected against. If the population has been settled in an environment where the relevant selective pressure exists, the learned behaviour is likely to be incorporated in the genome as fixed structure. (Example: How can Camels develop an innate ability to swim?)

In spite of the general acceptance of something like the baldwin effect, general scepticism prevails as to whether this effect can explain specific principles in UG (e.g. Projection Principle, the Principle of Full Interpretation etc.). Author refutes criticisms by offering a simulation of the evolution for the Principles-and-Parameters system (Chomsky, 1981).

Basic idea: Operate on a set of binary slots that can be filled in three ways:
- $0 =$ hardwired parameter setting („principle“), value 0
- $1 =$ hardwired parameter setting („principle“), value 1
- $? =$ open parameter, subject to learning (set to 0/1 during lifetime of individual)

A Fitness Algorithm is devised with takes the following steps:

1. Start with the genotype being evaluated
2. Choose another genotype from the population at random.
3. If the two genotypes are entirely fixed (contain no ?) and identical, then return a fitness of 2.
4. If the two genotypes match (see below), then attempt to establish communication between their owners. The fitness of the first genotype is proportional to how quickly the two genotypes learn to communicate. This fitness will be higher than 2.
5. Otherwise return fitness of 1.

Learning, in Turkel’s model, amounts to an algorithm which attempts to unify the values of the two genotypes by initializing the remaining ? in either genotype. After fitness had been calculated for each member of the population, reproduction took place following standard reproduction models in evolutionary biology resting on fitness of the agents.

The population consisted of genotypes which each possessed 12 relevant open slots (mimicking the full range of Princ plus Par in P&P) which were instantiated differently in each genotype (agent) in the first generation. (Different runs explored initial generations without plasticity = all principles; with equal plasticity and fixed values, and with more plasticity than fixed principles). The stable trend in the simulation was the following: In all except the no plasticity condition, each of the 12 P&P slots showed a clear trend to result in a fixed 1, a fixed 0 or a fixed ? in the long run. This is interpreted by the author such that evolution supports the emergence of a clear principle-and-parameter system with genetically fixed places for learning, and for UG. The trend was only driven by the benefit of successful communication (see fitness values). The trend of successful communication was not enough to evolve a homogenous all-principle community in the no-plasticity condition. The results with and without learning were substantially different.
Discussion and Critique:
While the essentials of the algorithm may be justly chosen for other types of genetic variation and selection, it is doubtful whether the paper operates within the basic assumptions of the Principle-and-Parameter theory, even though this theory is quoted as the core reference.

The Principles & Parameter dogma:
- Language acquisition from poor input is possible only because there are guiding principles that restrict the possible grammars in a way so as to direct the language learner towards any coherent natural language grammar.
- Hence, a principle can not be learned, and there can not be a class of natural languages that rest on more learned parameters for what later will turn into principles. In the absence of guiding parameters, children will acquire no natural language at all (and adults will not offer natural language input), including no principles, and no parameters either.
- A pre-grammar with no, or too few principles will hence at best enable its speakers to speak some agrammatical pre-language. This may come close to the Bickertonian development of language; in any case, the resulting development is exactly the one that the author set out to make superfluous. Language is predicted to arise gradually through different stages.
- If the author takes his position serious, he could maintain that no or too few principles result in no language at all. Then, however, he seems to miss the initial conditions for the Baldwin Effect. If humans can (in whatever way) learn the setting of a principle, but this acquisition does not bring them close to natural language, then they are not able to enter that evolutionary niece which would support the adoption of acquired parameters as fixed principles into the genome.

All in all, the article might be taken as a warning against too easy abuse of the convenient and binary talk of principles and parameters in UG.

3.6. Grammatical acquisition and linguistic selection


Computational simulation of a language learner („LAD“, language acquisition device) resting on the following components:
- A set of parameters for various syntactic features of a grammar, initially unset; potentially with a default setting that can be overruled; which is to be instantiated in reaction to trigger data.
- The fully specified parameter set will create one of a range of possible grammars defined by this „toy UG“.
- A statistically influenced learning procedure which implements a balance between ‘permanent parameter setting’ (i.e. a parameter once in place can never be removed) and ‘memoryless learning’ (a setting can be changed any time, in view of conflicting evidence; for example see Niyogi). The settings can be changed tentatively, or permanently, if enough supportive evidence arises.
- Triggering data: Sentence structures of one or other of the range of possible languages in the range of the toy UG. Trigger sets can be chosen so as to cohere with one grammar, ambiguously, or mixed to various proportions.

In the first series of runs, Briscoe explores the effect of default settings vs. no settings for the UG parameters. As is to be expected, default settings increase the speed of learning for languages which fit the defaults closely, and slows down learning for languages which do not fit the default; it never impeded learning altogether.
In a second series of runs, a number of agents, divided into producing adults, and learners, were collected in an evolutionary setup. With the usual interaction cycles, mating, birth and death of agents implemented, Briscoe studies the effects of default settings in mixed language communities, and their long term developments. Likewise, scenarios are implemented where the agents could accumulate more default parameter settings over time, and the development of such systems was evaluated.

Finally, Briscoe addresses the Creolization scenario. The kind of language acquisition device that is implemented in his setup is, inherently, incapable of innovation. Several types of mixed and impoverished input are tested in populations of agents; with mixed results.

3.6.1. The Language Acquisition Device

Classical Categorial Grammar as the structural backbone, including

- Forward Application
- Backward Application
- Forward Composition
- Backward Composition
- (Generalized Weak) Permutation.

(ex.s see p. 259)

Basic language families: Defined by unmarked order of verb V, subject S and objects O. Further specifications within each family: Order of modifiers and specifiers in phrases, order of adpositions, further phrasal orderings.

<table>
<thead>
<tr>
<th>Parameter s</th>
<th>Status</th>
<th>possible values</th>
<th>degree of evidence for given setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>CG options</td>
<td>application</td>
<td>A = absolute, principle</td>
<td>0 = rightward / 1 = leftward / unavailable</td>
</tr>
<tr>
<td></td>
<td>composition</td>
<td>D = default setting: initial value overridable by trigger</td>
<td>true available</td>
</tr>
<tr>
<td></td>
<td>permutation</td>
<td>? = unset</td>
<td>? = unset</td>
</tr>
<tr>
<td>grammatical categories</td>
<td>S</td>
<td>N</td>
<td>NP</td>
</tr>
</tbody>
</table>
| argorder | gendir | subjdir | objdir | ndir | prepdir | ...


Comment: The parameters are in part ordered in a lattice structure. E.g. the gendir parameter specifies the generic direction of functor-argument order in a given language to which one will refer in the absence of more specific information. More general parameters can be overruled by specific parameters.
The parser:
The system incorporates a CG parser which will parse new input on basis of the currently assumed
CG of the system. (parsing algorithm, see p. 263)

Parameter setting:
Statistical supported, n-local, partially ordered error-driven learning algorithm with limited
memory: Parameter resetting occurs if trigger input can not be parsed: n settings are flipped (at
random), if successful parse is achieved, the new settings get valued higher; a sequence of
confirmations will yield a permanent change. Partial ordering reflects the hierarchical relations
between some of the parameters.
Nontrivial statistical side conditions are needed in order to balance the probability values for the
parameters and turn them into a consistent, working CG for actual parses. (p. 265/66)

Implementation, p. 267 ff.

3.6.2. Acquisition from trigger by single learners
Unset learner: Forward, Backward Application, categories S and NP. Remaining parameters unset.
Default learner with additional settings for argorder, gendir, subjdir, v1, v2 = minimal SVO
right-branching grammar.
Both learners can change up to 4 parameters per trigger.

Trigger data from one of seven full languages in the grammar set (labels don’t imply a serious
statement about the NL grammar in question):
- 'English' (SVO, predominantly right-branching)
- 'Welsh' (SOVo1, mixed order)
- 'Malagasy' (VOS, right-branching)
- 'Tagalog' (VSO, right-branching)
- 'Japanese' (SOV, left-branching)
- 'German' (SOVv2, mixed-branching)
- 'Hixkaryana' (OVS, mixed branching)
- hypothetical OSV, left-branching

Triggers from target full language, uniformly-distributed fair sample, level-0 or level-1 sentences.

Result: Learning was reliably ensured by both learners, for any of the languages tested (as well as
a wide range of other possible full and subset languages in P-setting(UG)).

Speed of convergence depends on
default settings and target language
number of possible parameter changes per trigger
suitably chosen trigger set, and sequence of presentation
Optimal choice of learning setup leads to convergenvce already after low number of triggers.

Exploration: „Impure“ acquisition input: Ambiguous triggers that may likewise belong to a
different grammar.
150 triggers, randomly chosen from
SVO, -N, -PERM, -COMP plus SOVv2 („German“)
SVO, -N, -PERM, -COMP plus SOVo1 („English“)
(SVO, -N, -PERM, -COMP = intersection of „English“ and „German“ triggers)

Subset triggers conflict with target language in two parameters: objdir and v2, and argorder and
v1 respectively.
Bayesian acquisition algorithm is superior for conflicting and mixed input in comparison to non-statistical procedures, memoryless learning, deterministic parameter settings.

3.6.3. Learning in Populations

Language Agents (learner, parser, and generator) are integrated in Populations of Language Agents (32).

Interaction: randomly chosen agent generates utterance, directed towards a randomly chosen addressee Agent. Addressee attempts a parse.

Interaction cycle: Predefined number of interactions take place in the population.

Age of Agent: Defined in terms of interaction cycles; Age 1–4 count as learning age. Unparsable input is treated as trigger for acquisition. Age 10: Agent dies.

Reproduction: Above Age 4, random chosen agent pairs. Genetic transmission: Single point crossover, and single point mutation of the initial parameter settings. Reproduction can be made dependent on fitness of agent, which is measured in terms of successful communications.

(Simulation see fig. 9.8., page 279)

Simulation of development of varieties in mixed speaker communities (no genetic permutation; 32 default or unset agents.) Test for the relative „fitness“ of languages to survive in a mixed speaker community: 10 speakers of SVO, -N, -PERM, -COMP („English“ ∩ „German“) subset language plus 22 speakers of „German“ (SOVv2).

Result: Languages that were acquired b learners in this initial setup: SOVv2; SOVv2,-N; SVO; SVO,-N; SVO,-PERM,COMP. Yet, SOVv2 robustly survives in about 10 interaction cycles. Nevertheless, one can mimick the effect that speakers of other varieties turn up by imperfect learning from time to time.

Second test: Initial population of 16 speakers of SVO,-N,PERM,COMP („English“ ∩ „German“) subset language plus 16 speakers of „German“ (SOVv2) (50/50). Population fixed on SVO,-N,PERM,COMP in approximately 15 interaction cycles (both default learners and unset learners). This suggests that there are languages that are „easier“ to learn for the LAD, and survive in initial setups of equal distribution.

(Co-Evolution of the LAD and of language)

3.6.4. Creole scenarios

Populations of agents with subset languages only (= Pidgin stage). Behaviour of learners?
The LAD can only acquire, by positive evidence, the grammatical categories and their syntactic position, that are exemplified in the trigger input. The LAD is not able to spot missing expressive power of the trigger input, nor to invent new linguistic devices in order to fill in such gaps. Therefore, the system is unsuited to capture one strong innovative force in Creolization (as admitted in one footnote). Briscoe therefore concentrates on the process where language learners create word order regularities from mixed and inconclusive input.

- Pure impoverished ("Pidgin") input always resulted in reduced acquired languages, as the LAD can not innovate structure which is not exemplified.
- Impoverished input together with superstrate "English" input, together with an SVO default setting in syntax, yields SVO syntax.
- Other mixtures of impoverished input plus superstrate full input yield other, different results but mainly tend towards SVO word order in the long run, following the implemented default.

The default settings of parameters was in fact able to create a full grammar, following the SVO bias. It was only possible to achieve a full grammar with the additional influence of full superstrate input. This is empirically suspicious, because the abstract presentation of sentence structure neglects the fact that the superstrate input (and substrate, likewise) rests on a completely different lexicon and phonology. Bilingual language acquisition shows that children systematically keep input from different languages separate, performing a complete 1st language acquisition for each of the input languages. It would have to be shown that this universal ability breaks down completely just because one input variant is "defective" (which the acquiring child could certainly only spot over time, and not from the first trigger input on.). No known facts so far hint in this direction. The default settings of parameters were not able to mimick the rapid Creolization as described by Bickerton. While Bickerton maintains that Creolization can be performed within 2 generations, no matter what the input looks like, the simulations show a clear delay effect if the input parameters of the full language do not match the default settings (e.g. Berberice with superstrate Dutch). This is a further hint that the spontaneous ability to 'grammaticalize' a language if necessary is not adequately and fully simulated by default parameter settings in Briscoe’s LAD.

3.7. Conclusion.

While the combination of evolutionary models with linguistic developments is a challenging field, the true merits or limitations of the marriage will only become apparent once serious linguistic results inform the formal modelling.

Several of the models reviewed here show serious deficiencies, or fail to offer convincing models, because the open slots of evolutionary models seem to be filled with inappropriate, or inappropriately simple grammatical parameters. The influence of semantics and pragmatics on language evolution is barely recognized at all.

This is not to say that the marriage is doomed to fail. It is, however, a warning against taking even advanced grammatical theories and combining them with evolutionary processes by the sheer look of it. The change of English word order can offer an example.

In Part 1, we saw at length that some of the driving factors in play seemed to be pragmatic or due to parsing needs. Ambiguity results between a certain surface structure being discourse motivated or base grammar generated. Example: Optional right dislocation of 'heavy' constituents:

\[ \text{æt he [ ure sawle fram synna fagnyssum gehælan ] mæge} \]
\[ \text{,that he can heal our soul from the ulcers of sin'} \]
\[ \text{(from Fischer et al 2000: 155)} \]

\[ \text{æt hi mihton [ swa bealdlice Godes geleadan bodian ]} \]
\[ \text{,that they could preach God’s faith so boldly’} \]
\[ \text{(Fischer et al 2000: 156)} \]
We likewise saw that pragmatic information in language processing tends to be lost easily, as is known from processes like 'bleaching' (loss of presupposition). It is also suspected that pragmatic side informations are acquired later, and less reliably, by children than grammatical information in the core grammar (Guasty, 2002).

It was discussed at length that various formal approaches to syntax tend to cover such facts about language use, rather than point them out clearly. (This is understandable as long as linguists study such developments in exclusively syntactic theories rather than in integrated frameworks. Oddly, construction grammar at present seems the only integrated format that devotes itself to diachronic study. Other formats, notably the largescale integrated computational formats like HPSG or LFG which would be highly suited for integrated research, fail to show interest in language change.)

Clearly, if a formal theory of word order change in English takes such syntax theories as starting point without further reference to other modules of language, it is likely to miss essential driving forces in change. The resulting approach may still offer some kind of simulation of change, and even one where one can toy with parameters like probability distribution for possible utterances to mimick certain pragmatic effects, but then the true development is superimposed onto rather than derived from a given formal model.

On the other hand, nothing would stand against attempts to apply formal models of language evolution to a different set of parameters of a given language stage. (***) The preciseness of the results is tempting enough to cope with the initial complication of these systems.


A final word about the interdisciplinary enterprise of investigating language evolution in the „talking ape“ sense.

The investigation of language evolution requires the interaction of linguistics with biology, psychology, palaeo-anthropology, neurology, and primatology, at least.

Linguists need to rely on the other disciplines:
- Time scale: internal reconstruction, comparative method go back at best 6000 years. If Proto-Language coincides with the emergence of the first homo sapiens, it goes back 200 000 years, with leaves a gap of 194 000 years unaccountable by linguistic reconstruction methods. (Even if we date the invention of language back to earliest known stone age cultures, app. 40 000 years, this leaves us a gap of 34 000 dark years.)
- Long term equivalence of languages: In time distances like these, there are no unidirectional trends that would allow conclusions about the age of a language or its culture. The variation among languages, if it exists, levels out too quickly to allow for reconstruction. (See Part 2 on Typology, Creolization. The time spans in question are measured in centuries, not in millenia. See Newmeyer 2003; in Christiansen and Kirby 2003:58-76)

The non-linguistic disciplines need to rely on linguists:
- Communication is possible without syntax.
- Child language is not the same as proto language. In language acquisition, phylogensis does not repeat ontogenesis. (ftn. p. 81)
- Syntax is more than word order regularities.
- Natural language acquisition is not possible without principles of UG, hence principles of UG can not be acquired as part of language acquisition. (Bickerton shoots a juicy remark against the work of Nowak in this sense: p. 93)
- Language acquisition by adults is more likely to share the qualities of Proto-Language than child language.
Speakers of Pidgins witness that absence of grammar per se is not a linguistic feature that leads immediately to extinction of the community due to failure to communicate. (Although no speaker of Pidgin is without a second, full natural language, with may weaken the observation somewhat.)

None can operate without taking the others into account. E.g. purely socially motivated explanations of language evolving:

- **Grooming substitute** (Dunbar, in Christiansen and Kirby, 2003) — Linguistics: why then should language invariably convey factual information? Compare non-verbal grooming vocalization.
- **Gossip** — implausible with the initial small vocabularies.
- **Chance to enhance mating possibilities** — Biology: typically, features relevant for mating are developed in puberty, exactly when language acquisition ceases.
- **Setting up menstruation rituals for female bonding?** — …
- **Yo-di-doo theory: First vocalizations were made to facilitate and coordinate common labour?**
- **Initiating marriages, so men would not know who they weren’t supposed to cheat with?** — Again, abilities that are selectively relevant only in adulthood undergo selective pressures that will not move the critical developmental period between age 0 and 6. For the talking male to successfully control his offspring, it would be fully sufficient to talk at age 12 - 15. For females, the language faculty would be completely non-selective.

Division of language faculty into three distinct subsystems:

(a) **Modality**: Development of the vocal tract. (Unlike earlier assumptions, the low larynx is not confined to humans; Fitch, 2005. Good vocal abilities, e.g. in parrots, come independent of language faculty. Gesturing and sign language are independent viable communication systems.)

(b) **Symbolic abilities**: Higher animals master symbolic reference, naming. (Latest research (Fischer, Tomasello): dog vocabulary learning with human like acquisition strategies.) Bickerton: Selected by advantage of cooperating in changed environment.

(c) **Structure**: Mastery of structured composition system for strings of symbolic expressions; „syntax“. Communication without structure is possible (telegarms, headlines, Pidgins, etc.); structure without communication? Bickerton: Symbolism may have preceded syntax by as much as 2 Million years.

Holism, Synthetism, and Syntax:

Bickerton defends the view that synthetic language structure was present from the first stage on that he would be willing to call language: Importance to take synthetic (proto-compositional?) language structure seriously and to distinguish between holistic alarm call, threat, fear - type calls („better conveyed by body language anyway“) pre-syntactic but synthetic proto language utterances full syntax

„One hears frequently of „proto-syntax‘, which seems to mean one-clause sentences with fixed word order, and there is a widespread but wholly erroneous belief that this does not merely constitute a step in the direction of real syntax, but that once one has achieved such a level of structure, real syntax follows automatically.“ (p. 87)

Regular word order ≠ Real syntax ≠ P&P in the original form ≠ Bickerton’s surface minimalism?

The quest for the unified field theory for language remains a challenge …
References:


K.E. Georges: Lateinisch-Deutsch: Ausführliches Handwörterbuch (Digitale Ausgabe)


Trudgill, Peter. 1989. Contact and evolution in linguistic change. In: Leiv Egil Breivik & Ernst
Håkon Jahr (eds.): Language Change: Contribution to the study of its cause, 227-237.
Berlin, de Gruyter.
254.
T.a. in: xx
Available at the author's homepage.
and Quantification. Dordrecht, Kluwer: 177-238.