Diachronic Generative Syntax: from 2 Factors to 3

Theresa Biberauer

University of Cambridge, Stellenbosch University and CRISSP (KU Leuven)

Clarification

- Our starting point: clarifying the factors

(1) The classic generative "algorithm"

Universal Grammar (UG) + Input (PLD) ➔ Adult Grammar (I-language)

rich UG

2 factors

(2) The minimalist generative "algorithm"

UG + PLD + general cognitive factors ➔ Adult Grammar (I-language)

impoverished UG > language structure = largely **emergent**

3 factors

The plan

Session 1 (Wednesday 24 June)

Diachronic generative syntax in the 2 and pre-3 Factors era (1979 - ±2010)

- an **acquisition**-centred approach to variation and change
- a **parameter**-centred approach: 'classic' and post-1980s parameters
- problems with UG-given (Factor 1-based) parameters
- **pathways of change**: grammaticalisation and cycles

throughout: "lurking" precursors to Factor 3

Session 2 (Thursday 25 June)

Diachronic generative syntax in the 3 Factors era (since ca. 2010)

- an **emergentist** approach to acquisition, variation and change: the Maximise Minimal Means (MMM) model
  - a role for children and for adults
- emergent parameters and parametric change
- grammaticalisation, pragmaticalisation, cycles
- exaptation
- I-contact: simplification and complexification (the relevance of sociosyntax)

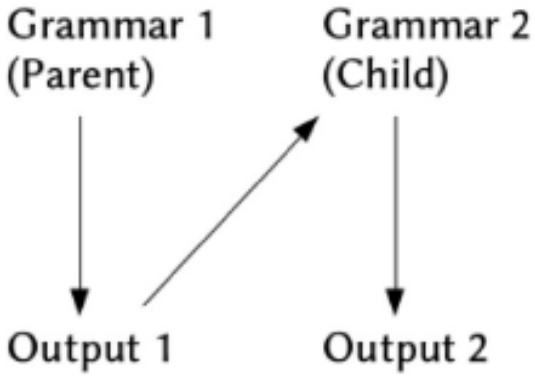
Generative diachronic syntax: preliminaries

- The centrality of the **(first) language-acquiring child**
  - Most change happens in childhood (Weinreich, Labov & Herzog 1968: 104, Andersen 1973)
- A generative paradox:
  - children = exceptional language learners ('little inflection machines', etc.)
    - ➢ flawless transmission; no change
  - grammars change:

Generative diachronic syntax: preliminaries

- The centrality of the **(first) language-acquiring child**
  - Most change happens in childhood (Weinreich, Labov & Herzog 1968: 104, Andersen 1973)
- A generative paradox:
  - children = exceptional language learners ('little inflection machines', etc.)
    - ➢ flawless transmission; no change
  - grammars change:

(3)



"Z-model" of language change

Source: Andersen (1973: 767)

- no direct access to Grammar 1
- no access to **negative evidence** (direct or indirect)
- acquisition proceeds via **abduction**

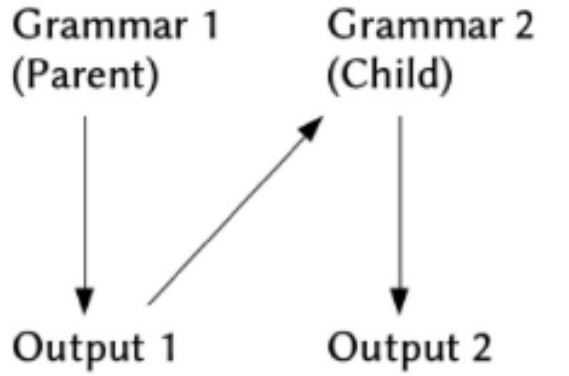
Generative diachronic syntax: preliminaries

- What is **abduction**? (Andersen 1973, Chomsky 1968: 90ff.; but NB Deutscher 2002, Walken 2011)

(4) 'Change caused by the fact that learners only have access to the output of a generative grammar … and to Universal Grammar … with no direct access to the grammar itself. The combination of primary linguistic data … and Universal Grammar may lead the learner to abduce a system which is distinct from that underlying the primary linguistic data by reanalysis …' (Roberts 2007: 445)

- This is actually <u>just</u> **reanalysis** (Timberlake 1977, Langacker 1977, Harris & Campbell 1995):

  change to an underlying structure with no immediate surface change (≈ **actuation**)



"Z-model" of language change

(5) a. [It is bet **for me**] [te sleen my self than ben defouled thus.]

'It is better for me to slay myself than to be violated thus.' (Chaucer, *Franklin's Tale)*

b. [It is bet] [**for me** te sleen my self than ben defouled thus.]

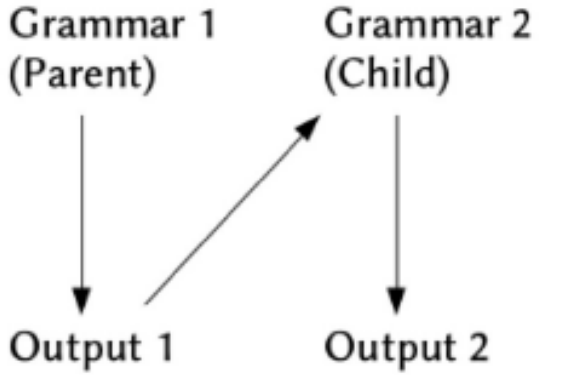
c. [**For me** to slay myself] [would be better than to be violated thus.] (≈ **actualisation**)

2-Factor generative diachronic syntax

- Back to the **2 factors**

(6) 'In a highly idealized picture of language acquisition, UG is taken to be a characterization of the child's pre-linguistic initial state. Experience – in part, a construct based on an internal state given or already attained – serves to fix the parameters of UG, providing a core grammar, guided perhaps by a structure of preferences and implicational relations among the parameters of the core theory.' (Chomsky 1981: 7)

  - Factor 1: UG "steering" acquisition
  - Factor 2: Input ≠ 'everything the child hears'; = **intake** (Evers & van Kampen 2008, Gagliardi 2012, 2013, Lidz & Gagliardi 2015))



"Z-model" of language change

- 2 diachronically relevant notions:

  - the distinction between the **core** and the **periphery** of the grammar (p.8)

  - the role of the 'ideal speaker-hearer'

2-Factor generative diachronic syntax

- **Core**: main-clause *wh*-question formation

  (7) a. [What] **will** the reporters find?

  b. [Where] **have** the team left their kit?

- **Periphery**: Locative Inversion (8) and Quotative Inversion (9)

  (8) a. [Down the hill] **rolled** the rock.

  b.[On the table] **stands** a lamp.

  (9) a. [‘What is this about?’] **wondered** the students/*they.

  b. [‘That’s a crazy idea!’] **said** Alex/?(s)he.

And where would (10) fit?

(10) a. **Man** is that annoying!!

b**. Heck** can she sing!

2-Factor generative diachronic syntax

- **Core**: Degree Adverb – Adjective

  (11) **so/too/as/sufficiently** big

- **Periphery**: Adjective – Degree Adverb

  (12) big **enough**

  and

- **Core**: Degree-Adjective – Noun

  (13) **much/more/sufficient/enough** pudding

- **Periphery**: Noun – Adjective

  (14) pudding **enough/aplenty/galore**

2-Factor generative diachronic syntax

- **Core**: Degree Adverb – Adjective

(11) **so/too/as/sufficiently** big

- **Periphery**: Adjective – Degree Adverb

(12) big **enough**

and

- **Core**: Degree-Adjective – Noun

(13) **much/more/sufficient/enough** pudding

- **Periphery**: Noun – Adjective

(14) pudding **enough/aplenty/galore** (see Yang 2016a,b on the *a*-forms)

2-Factor generative diachronic syntax

- What about the **'ideal speaker-hearer**'?

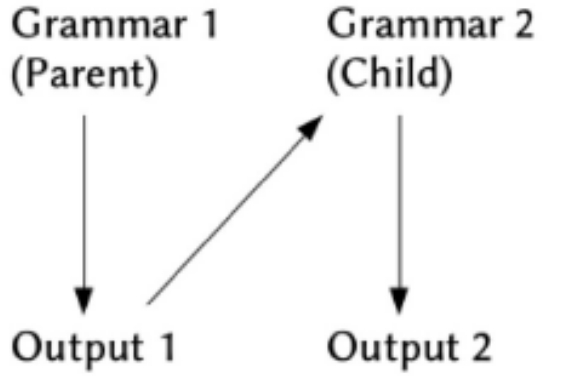
Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance. (Chomsky 1965: 3)

- From the perspective of diachronic generative syntax, something does need to change.
- In terms of the Z-model, there are 2 possibilities:

(i) the nature of the input (= Output 1)

- structural ambiguity, change in (surface/E-language) distributions

(ii) the nature of the acquirer (bilingual, L2, etc.)



"Z-model" of language change

- (i) creates a chicken-and-egg paradox
- (ii) limits change to 'contact' scenarios ('perfect' vs 'imperfect' learning)

2-Factor generative diachronic syntax

- **Structural ambiguity**: recall the **reanalysis** scenario we've encountered:

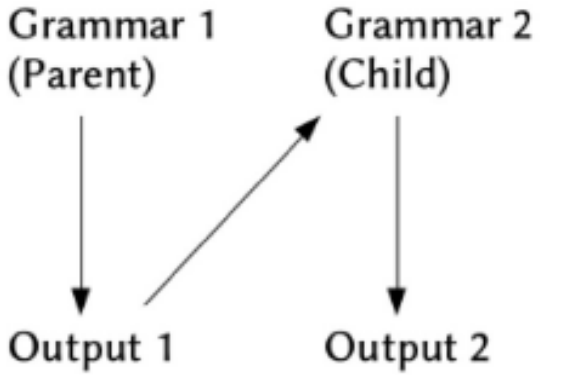
(5) a. [It is bet **for me**] [te sleen my self than ben defouled thus.]

'It is better for me to slay myself than to be violated thus.' (Chaucer, *Franklin's Tale)*

b. [It is bet] [**for me** te sleen my self than ben defouled thus.]

c. [**For me** to slay myself] [would be better than to be violated thus.] (≈ **actualisation**)

- The grammatical analyses in (5a) and (5b) could define 2 distinct grammars:
  - Grammar 1: (5a)
  - Grammar 2: (5b)
- ➢ Why don't Grammar 2-speakers start using (5c) right away?
- ➢ What would originally have triggered the postulation of (5b)? The parent generation were fine …



"Z-model" of language change

2-Factor generative diachronic syntax

- **Distributional change in E-language**:

  'Sometimes … minor changes in the relevant childhood experience cross a threshold and have consequences for the grammars which emerge. In that case, we have a different grammar, and this may entail far-reaching surface effects in terms of the class of sentences generated by the new grammar.' (Lightfoot 1999: 79)

  ➢ Can this help us to understand attested changes like OV > VO (15a) and the loss of fuller V2 (15b) in English?

  (15) a. Old English: … that the students **the books** **read**. > PDE: … that the students **read** **the books**.

  b. Yesterday **read** the students the books. > PDE: Yesterday the students **read** the books.

  - ambiguous input: e.g. SVO-structures (ditto SV-structures)

  (16) a. The students read the books.

  b. $[_{CP}$ The students$_i$ read$_x$-C $[_{TP}$ t_i t_x $[_{VP}$ the books $t_x]]]$ -- V2 and OV ("German"-style V2)

  c. $[_{CP}$ The students$_i$ read$_x$-C $[_{TP}$ t_i t_x $[_{VP}$ t_x the books$]]]$ -- V2 and VO ("Scandinavian"-style V2)

  d. $[_{CP}$ $[_{TP}$ The students read$_x$-T $[_{VP}$ t_x the books$]]]$ -- V-to-T and VO ("French"-style VO)

  e. $[_{CP}$ $[_{TP}$ The students T $[_{VP}$ read the books$]]]$ -- V-in-V and VO ("PDE"-style VO)

2-Factor generative diachronic syntax

- **Distributional change in E-language**:
  - ambiguous input: e.g. SVO-structures (ditto SV-structures)

(16) a. The students read the books.

b. $[_{CP}$ The students$_i$ read$_x$-C $[_{TP}$ t_i t_x $[_{VP}$ the books $t_x]]]$ -- V2 and OV ("German"-style V2)

c. $[_{CP}$ The students$_i$ read$_x$-C $[_{TP}$ t_i t_x $[_{VP}$ t_x the books$]]]$ -- V2 and VO ("Scandinavian"-style V2)

d. $[_{CP}$ $[_{TP}$ The students read$_x$-T $[_{VP}$ t_x the books$]]]$ -- V-to-T and VO ("French"-style VO)

e. $[_{CP}$ $[_{TP}$ The students T $[_{VP}$ read the books$]]]$ -- V-in-V and VO ("PDE"-style VO)

- For ambiguous input to matter, it needs to "count" more than it previously did for the parent generation.
  - lesser use of unambiguous structures (e.g. O-V-S and Adv-V-S for V2; "30-%" YP-V-S)
  - What would produce this distributional change, if not a prior I-language change?
  - particularly hard to understand in 'classic' Principles & Parameters period: parameters = 'macro'

2-Factor generative diachronic syntax

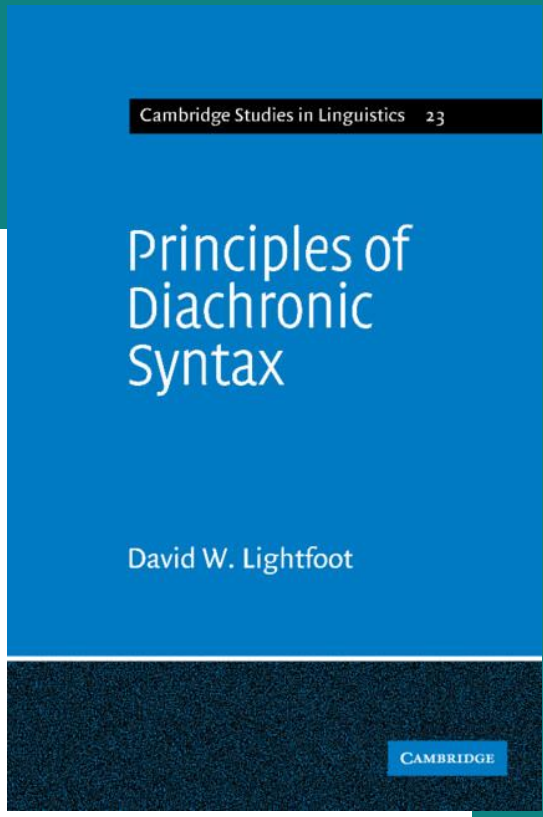
- From the perspective of diachronic generative syntax, something does need to change.
- In terms of the Z-model, there are 2 possibilities:

  (i) the nature of the input (= Output 1)

  - structural ambiguity, change in (surface/E-language) distributions

  (ii) the nature of the acquirer (bilingual, L2, etc.)

2-Factor generative diachronic syntax



- From the perspective of diachronic generative syntax, something does need to change.
- In terms of the Z-model, there are 2 possibilities:

(i) the nature of the input (= Output 1)

- structural ambiguity, change in (surface/E-language) distributions

(ii) the nature of the acquirer (bilingual, L2, etc.)

- From the earliest work:

(17) The **Transparency Principle**:

'[The Transparency Principle] requires derivations to be minimally complex and initial, underlying structures to be 'close' to their respective surface structures' (Lightfoot 1979: 121).

'If the Transparency Principle characterizes the limits to the permitted degree of exceptionality or derivational complexity, then it will follow from this principle that as these limits are approached so some kind of therapeutic reanalysis will be necessary to eliminate the offending complexity' (Lightfoot 1979: 122).

e.g. the Old English modals *cunnan, magan, motan, sculan* and *willan* with their growing list of 'exception features' (preterite-presents, no –*s*, past-forms without past meanings, no non-finite forms)

➢ weird class of Vs > class of I-elements (reanalysis)

2-Factor generative diachronic syntax

- The Transparency Principle was initially understood as a **condition on derivations**.
- In the 2 Factors (classic GB/Principles & Parameters) era, it could also be thought of as a condition on parameter setting, i.e. part of Factor 1.
- Lightfoot (1991): **Degree-Zero learnability** (see also Wexler & Culicover's 1980 'Degree-2 learnability')
  - *degree* = extent of clausal embedding; *degree-zero* = main-clause only
    - Recall our V2/SOV/SVO example (16): 'if parameter setting were sensitive to embedded material, there would have been plenty of robust data to warrant object-verb order'

- where V raises to C (=V2), we don't have unambiguous evidence for VP directionality:

(18) a. $[_{CP}$ The students$_i$ **read$_x$**-C $[_{TP}$ t_i t_x $[_{VP}$ the books **t_x**$]]]$ -- V2 and OV ("German"-style V2)

b. $[_{CP}$ The students$_i$ **read$_x$**-C $[_{TP}$ t_i t_x $[_{VP}$ **t_x** the books$]]]$ -- V2 and VO ("Scandinavian"-style V2)

- where C is filled with a complementiser, V unambiguously remains in its base position:

c. . $[_{CP}$ … that C $[_{TP}$ the students T $[_{VP}$ the books **read**$]]]$

2-Factor generative diachronic syntax

- So is Degree-Zero learnability part of UG (Factor 1)?

  Lightfoot (1991: 40): 'the claim that children are degree-zero learners … reflects a property of their "learning" capacity and not of Universal Grammar'

  - ➢ a foreshadowing of Factor 3?

- Some important **predictions**:

  - main clauses will 'lead' change
  - changes in embedded clauses = 'an automatic by-product of certain changes in unembedded domains' (1991: 43)
    - ➢ 'catastrophic' change in embedded clauses, to some extent in contrast to main clauses (shifting distributions)
      - Ross's (1973) **Penthouse Principle** seems to hold universally (though see Andersson 1974 on Swedish *ha*-deletion, which had a short-lived equivalent in earlier German; see Breitbarth 2005)

A closer look at UG (Factor 1)

- The 'classic' Principles & Parameters perspective:
  - invariant principles (e.g. structure dependence)
  - variable (binary) parameters, requiring setting via the PLD

(19) a. **Head Parameter**:

The head (X) of a phrase (XP) PRECEDES/FOLLOWS its complement.

b. **V2 Parameter:**

The finite verb MOVES/DOESN'T MOVE to C, the head of CP.

c. **V-to-T Parameter:**

The finite verb MOVES/DOESN'T MOVE to T, the head of TP.

d. **Wh-movement Parameter:**

The wh-phrase MOVES/DOESN'T MOVE to Spec-CP.

e. **Null Subject Parameter:**

Pronoun subjects CAN/CANNOT remain unrealised in finite clauses.

A closer look at UG (Factor 1)

- The 'classic' Principles & Parameters perspective:
  - invariant principles (e.g. structure dependence)
  - variable (binary) parameters, requiring setting via the PLD

(19) a. **Head Parameter**:

The head (X) of a phrase (XP) PRECEDES/FOLLOWS its complement.

b. **V2 Parameter:**

The finite verb MOVES/DOESN'T MOVE to C, the head of CP.

c. **V-to-T Parameter:**

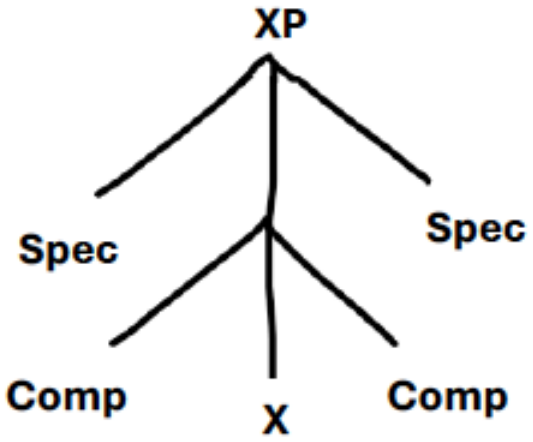
The finite verb MOVES/DOESN'T MOVE to T, the head of TP.

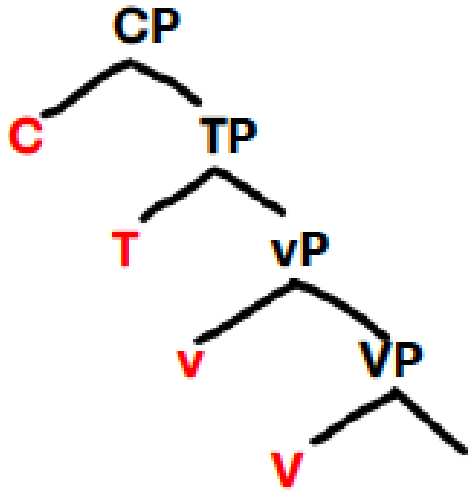
d. **Wh-movement Parameter:**

The wh-phrase MOVES/DOESN'T MOVE to Spec-CP.

e. **Null Subject Parameter:**

Pronoun subjects CAN/CANNOT remain unrealised in finite clauses.





- Initially: **parameterised principles**

A closer look at UG (Factor 1)

- From a diachronic perspective:
  - change = parameter **resetting** (linguists' perspective)

    e.g. loss of OV, (fuller) V2 and V-to-T in the history of English

(19) a. **Head Parameter**:

The head (X) of a phrase (XP) PRECEDES/**FOLLOWS** its complement.

b. **V2 Parameter:**

The finite verb **MOVES**/DOESN'T MOVE to C, the head of CP.

c. **V-to-T Parameter:**

The finite verb **MOVES**/DOESN'T MOVE to T, the head of TP.

A closer look at UG (Factor 1)

- From a diachronic perspective:
  - change = parameter **resetting** (linguists' perspective)

    e.g. loss of OV, (fuller) V2 and V-to-T in the history of English

(19) a. **Head Parameter**:

The head (X) of a phrase (XP) **PRECEDES**/FOLLOWS its complement.

b. **V2 Parameter:**

The finite verb MOVES/**DOESN'T MOVE** to C, the head of CP.

c. **V-to-T Parameter:**

The finite verb MOVES/**DOESN'T MOVE** to T, the head of TP.

Recognition for the Head Parameter:
it can't just be a binary option;
category-specific setting must be possible

A closer look at UG (Factor 1)

- From a diachronic perspective:
  - change = parameter **resetting** (linguists' perspective)

    e.g. loss of OV, (fuller) V2 and V-to-T in the history of English

(19) a. **Head Parameter**:

The head (X) of a phrase (XP) **PRECEDES**/FOLLOWS its complement.

b. **V2 Parameter:**

The finite verb MOVES/**DOESN'T MOVE** to C, the head of CP.

c. **V-to-T Parameter:**

The finite verb MOVES/**DOESN'T MOVE** to T, the head of TP.

Recognition for the Head Parameter:
it can't just be a binary option;
category-specific setting must be possible

- BUT: the classic parameters were coarse-grained > **catastrophic** changes, where change *seems* **gradual**

  e.g. co-occurring in the output of a single speaker:

(20) a. The students have **the books read**. [OV]

b. The students have **read the books**. [VO]

A closer look at UG (Factor 1)

- Competing grammars (Santorini 1989, 1992, Kroch 1989, Taylor 1990, Pintzuk 1991, etc.)
  - Where the **input** requires it, children do not have to set a given parameter uniquely.

e.g. OV and VO co-occurring in the output of a single speaker:

(20) a. The students have **the books read**. [OV]

b. The students have **read the books**. [VO]

➢ A 'double base':

(19) a. **Head Parameter**:

The head (X) of a phrase (XP) **PRECEDES**/**FOLLOWS** its complement.

A closer look at UG (Factor 1)

- **Competing grammars** (Santorini 1989, 1992, Kroch 1989, Taylor 1990, Pintzuk 1991, etc.)
  - Where the **input** requires it, children do not have to set a given parameter uniquely.

e.g. OV and VO co-occurring in the output of a single speaker:

(20) a. The students have **the books read**. [OV]

b. The students have **read the books**. [VO]

➢ A 'double base':

(19) a. **Head Parameter**:

The head (X) of a phrase (XP) **PRECEDES**/**FOLLOWS** its complement.

- must hold for X = V and X = T

(21) a. … that the students the books **read have** [OV-Aux]

b. … that the students **have read** the books. [Aux-VO]

c. … that the students **have** the books **read**. [Aux-OV]

But:
d. **… that the students **read** the books **have***
is unattested (Fuss & Trips 2002, Biberauer et al. 2014)
➢ the **Final-over-Final Condition** gap

A closer look at UG (Factor 1)

- Competing grammars applied to classic parameters:
  - significantly complicate the assumptions about child-language acquisition > increased hypothesis space
    - with n=30 binary parameters, the number of possible grammars is 2^{30} = 1 073 741 824
    - if some/all of these parameters can be "set twice" …
  - over-generate
  - don't fit with what is known about **doublets** (Kroch 1994)

    > shouldn't be a long(er)-term phenomenon (e.g. stable variation shouldn't be a possible outcome)
  - don't obviously predict the **directionality** of the change (always a challenge, given acquisitional discontinuity):

    e.g. VO and head-initiality more generally win out in English > Why?

Detour: The interaction between UG and the input

- Recall: UG is assumed to "steer" the acquirer in interpreting the input.
- **Cues/triggers/p(arameter)-expressions** = key (Dresher & Kaye 1990, Dresher 1999; Lightfoot 1999, 2006, Lightfoot & Westergaard 2007, Westergaard 2017)

(22) Quantity (in)sensitivity (stress placement; Dresher 1999: 31)

a. Parameter: The language **DOES/DOES NOT** distinguish between light and heavy syllables.

b. Default: Assume all syllables have the same status (QI).

c. Cue: Words of *n* syllables, conflicting stress contours (QS).

- NB: the default is "preset" in UG and doesn't require special input.

Detour: The interaction between UG and the input

- Recall: UG is assumed to “steer” the acquirer in interpreting the input.

- In syntax, cues correspond to abstract pieces of structure.

Detour: The interaction between UG and the input

- Recall: UG is assumed to "steer" the acquirer in interpreting the input.

- In syntax, cues correspond to abstract pieces of structure.

(23)

a.	$[_{CP}$ XP V-C $[_{TP}$ Subject …]	--- V2
b.	$[_{TP}$ Subject V-T … Adv/*not* $[_{VP}$ t_V Object]	--- V-to-T
c.	$[_{TP}$ Subject T … Adv/*not* $[_{VP}$ V Object]	--- V in situ
d.	$[_{VP}$ V Object]	-- VO (head-initial VP)
e.	$[_{VP}$ Object V]	-- OV (head-final VP)

  - Some questions:
    - Where are the cues? (Lightfoot 1999 vs the rest)
    - How do acquirers use them?

Detour: The interaction between UG and the input

- Cues are assumed to facilitate acquisition by enabling the acquirer to link up components of UG (typically, the unset parameters) with the input.

- Using the Clark & Roberts (1992) notion of *p-expression* ≈ cue (see also Roberts 2021)

(24) **P-expression** (≈ cue): A substring of the input text *S* expresses a parameter p_i just in case a grammar must have p_i set to a definite value in order to assign a well-formed representation to S.

➢ cues/p-expressions = (in some sense) the component of the input (Factor 2) that UG 'steers' the acquirer to.

- For classic cue-based approaches, the input leads to fixing (setting) of the relevant parameter.

- In Yang's (2002) **Variational Learning** model:

(25) Upon the presentation of an input datum (=substring) *s*, the child:

a. selects a grammar G_i with the probability p_i, and

b. analyses *s* with G_i

c. if successful, reward G_i by increasing p_i; otherwise, punish G_i by decreasing p_i.

Detour: The interaction between UG and the input

- So far, we have assumed unambiguous cues. All input strings aren't equally informative, however.
  - Recall:

(18) a. $[_{CP}$ The students$_i$ **read$_x$**-C $[_{TP}$ t_i t_x $[_{VP}$ the books **t_x**]]] -- V2 and OV ("German"-style V2)

b. $[_{CP}$ The students$_i$ **read$_x$**-C $[_{TP}$ t_i t_x $[_{VP}$ **t_x** the books]]] -- V2 and VO ("Scandinavian"-style V2)

- On Yang's model, ambiguous strings will lead to a reward for **both** the VO and the OV grammar.
- Back to Competing Grammars and the directionality-of-change question:

➢ an intriguing possibility: differences in the 'fitness' of the competing options in generating the input (see Heycock & Wallenberg 2013 for discussion of the loss of V-to-T in Faroese and Mainland Scandinavian, where the V in situ grammar emerged as the 'fittest' of the competing options)

[Note: Generative approaches differ in how they model the role of ambiguous input. Consider Roberts (2021) on **p-ambiguity**:

a. **P-ambiguity**: A substring of the input text S is strongly p-ambiguous with respect to a parameter p_i just in case a grammar can have p_i set to either value and assign a well-formed representation to S.

b. A strongly p-ambiguous string may express either value of p_i and therefore **trigger either value of p_i**. (my highlighting and emphasis; TB]

A closer look at UG (Factor 1)

End of Detour/Back to Competing Grammars:

- "Classic" competing grammars can be seen as a response to the need to model variation with unhelpfully large UG-given parameters
  - Contrast how we model the availability of both VO and OV orders in OV V2 Germanic languages.
- Subsequent empirical and theoretical developments suggest that capturing the OV-VO variation observed in Old and Middle English requires more fine-grained parameters.
  - From the perspective of credibly modelling change across successive generations of speakers, we would expect the differences in formal encoding to be minimal (micro not macro).
  - Parameter-setting is necessarily "catastrophic", whereas change *seems* gradual; if the parametric specifications are fine-grained enough, though, we may the *impression* of gradualness, where this is not in fact the case.
  - And this is particularly so if we remember the reality that real speakers/hearers differ from their ideal counterparts in (partially) controlling multiple registers, which they may not all have acquired as kids. They are 'multilingual' in this sense.
  - ➢ The question (for discussion): do Competing Grammars help us to model this finer-grained variation?
    - To be borne in mind:
      - the need to pre-specify the options that compete (see Heycock & Wallenberg 2013: 145, note 26 on this)
      - the extent to which alternate forms in grammars are acquired simultaneously (consider registers) > do we want competing grammars or maybe more of a layered grammar model?

[See Wallenberg et al. 2026 for a positive evaluation of the Competing Grammars Model, particularly when harnessed within Yang's Variational Learning paradigm]

A closer look at UG (Factor 1)

- A little more on the finer-grained parameters of the post-classic UG-given parameters period:
  - both **empirical** and **theoretical** developments led to increased granularity in the formulation of parameters
- An empirical development: increased knowledge of the typology of OV and VO systems

(26) A typology of OV systems (Biberauer & Sheehan 2013):

a. "rigid" head-finality: Japanese, Malayalam, etc.

b. clausal head-finality, nominal head-initiality

c. "leaking" OV of different kinds, e.g. West Germanic

d. OVX, where O is the direct object (Baker 2005, Hawkins 2009)

e. $O_{[F]}VX$, where $O_{[F]}$ is a restricted object-type (e.g. Neg, Focused, pronominal, etc.)

f. more lexically restricted OV (e.g. only with certain predicates)

A closer look at UG (Factor 1)

- How knowledge of the typology of OV systems matters to diachronically oriented generativists:

(26) A typology of OV systems:

a. "rigid" head-finality: Japanese, Malayalam, etc.

b. clausal head-finality, nominal head-initiality

c. "leaking" OV of different kinds, e.g. West Germanic

d. OVX, where O is the direct object (Baker 2005, Hawkins 2009)

e. $O_{[F]}VX$, where $O_{[F]}$ is a restricted object-type (e.g. Neg, Focused, pronominal, etc.)

f. more lexically restricted OV (e.g. only with certain predicates)

➢more and less 'macro' OV patterns

A closer look at UG (Factor 1)

- How knowledge of the typology of OV systems matters to diachronically oriented generativists:

(26) A typology of OV systems:

a. "rigid" head-finality: Japanese, Malayalam, etc.

b. clausal head-finality, nominal head-initiality

c. "leaking" OV of different kinds, e.g. West Germanic

- CP extraposition (… *said has* **[$_{CP}$ *that* …]**; optional PP extraposition (… *the book* (***in the library***) *read* (***in the library***))

d. OVX, where O is the direct object (Baker 2005, Hawkins 2009)

e. $O_{[F]}VX$, where $O_{[F]}$ is a restricted object-type (e.g. Neg, Focused, pronominal, etc.)

f. more lexically restricted OV (e.g. only with certain predicates)

A closer look at UG (Factor 1)

- How knowledge of the typology of OV systems matters to diachronically oriented generativists:

(26) A typology of OV systems:

a. "rigid" head-finality: Japanese, Malayalam, etc.

b. clausal head-finality, nominal head-initiality

c. "leaking" OV of different kinds, e.g. West Germanic

- CP extraposition (… *said has* **[$_{CP}$ *that* …]**; optional PP extraposition (… *the book* (***in the library***) *read* (***in the library***))

d. OVX, where O is the direct object (Baker 2005, Hawkins 2009)

e. $O_{[F]}VX$, where $O_{[F]}$ is a restricted object-type (e.g. Neg, Focused, pronominal, etc.)

f. more lexically restricted OV (e.g. only with certain predicates)

A closer look at UG (Factor 1)

- How knowledge of the typology of OV systems matters to diachronically oriented generativists:

(26) A typology of OV systems:

a. "rigid" head-finality: Japanese, Malayalam, etc.

b. clausal head-finality, nominal head-initiality

c. "leaking" OV of different kinds, e.g. West Germanic

- CP extraposition (… *said has* **[$_{CP}$ *that* …]**; optional PP extraposition (… *the book* (***in the library***) *read* (***in the library***))

d. OVX, where O is the direct object (Baker 2005, Hawkins 2009)

e. $O_{[F]}$VX, where $O_{[F]}$ is a restricted object-type (e.g. Neg, Focused, pronominal, etc.)

f. more lexically restricted OV (e.g. only with certain predicates)

A closer look at UG (Factor 1)

- How knowledge of the typology of OV systems matters to diachronically oriented generativists:

(26) A typology of OV systems:

a. "rigid" head-finality: Japanese, Malayalam, etc.

b. clausal head-finality, nominal head-initiality

c. "leaking" OV of different kinds, e.g. West Germanic

- CP extraposition (… *said has* **[$_{CP}$ *that* …]**; optional PP extraposition (… *the book* (***in the library***) *read* (***in the library***))

d. OVX, where O is the direct object (Baker 2005, Hawkins 2009)

e. $O_{[F]}$VX, where $O_{[F]}$ is a restricted object-type (e.g. Neg, Focused, pronominal, etc.)

f. more lexically restricted OV (e.g. only with certain predicates)

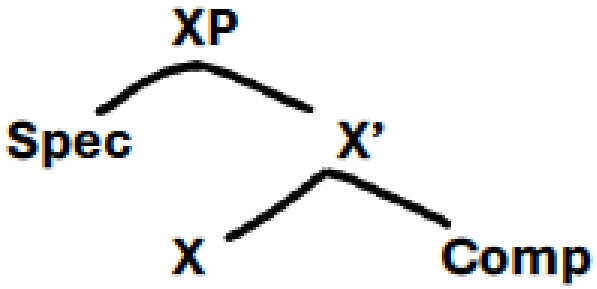
A closer look at UG (Factor 1)

- Theoretical developments that have impacted on the formulation of parameters:
  - Kayne (1994): Antisymmetry

A closer look at UG (Factor 1)

- Theoretical developments that have impacted on the formulation of parameters:
  - Kayne (1994): Antisymmetry

(27)



- All surface word orders derive from a Spec-Head-Comp configurations.
- Surface VO may or may not involve V (and other movement)
- Surface OV must involve some movement that places O to the left of V.

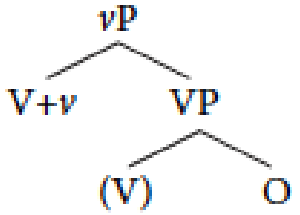
- rigid head-final order (e.g. in (26a) in the OV typology): 'roll-up' movement (Comp-to-Spec)
- OV orders where only the Object or a sub-type of Object precedes V (e.g. (26d) and (26e) in the OV typology): independent Object movement to a higher projection (e.g. Spec-AgrOP)
- West Germanic-style OV: short V-movement + Object-piedpiping

A closer look at UG (Factor 1)

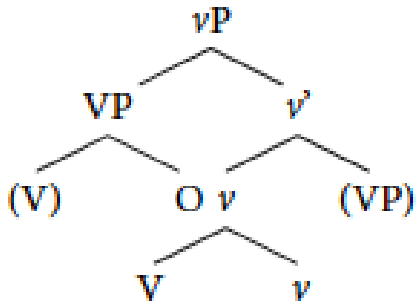
- West Germanic-style OV: short V-movement + Object-piedpiping (Biberauer & Roberts 2005 et seq.)

(28)

i. V-to-*v* raising:



ii. VP-to-(inner)Spec*v*P movement:



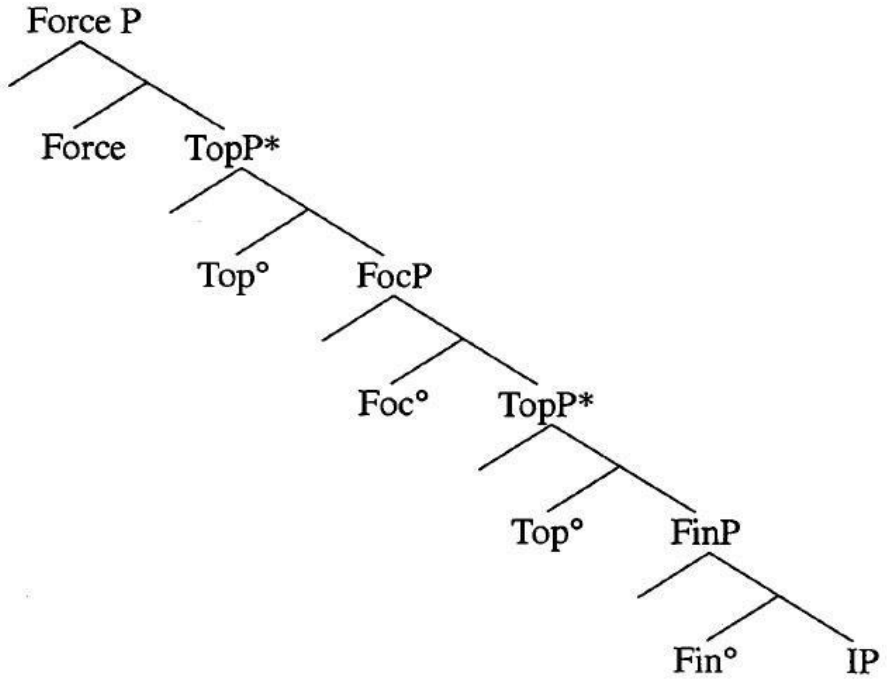
- Object **piedpiping** alternates with **stranding** (i.e. Object-only movement), as it does in other cases (e.g. *Which drawer did you leave your homework in?* vs *In which drawer did you leave your homework?*)
- Object-only movement produces OVX structures ('leaking')
- The loss of the piedpiping option produces a less rigid OV system, e.g. some kind of OVX (26d/e) system.
- Earlier English, like modern West Germanic, already had special Object movement operations (Scrambling), so information structure [topic/focus-related considerations] would have played into the kind of OVX system that arose in Middle English (see much work by Susan Pintzuk, Ann Taylor, Wim van der Wurff, Ans van Kemenade, Tara Struik, and others)

➢ This mode of analysis doesn't require a 'double base'/doubly set Head Parameter (no Competing Grammars)

A closer look at UG (Factor 1)

- Further theoretical developments that have impacted on the formulation of parameters:

(29)



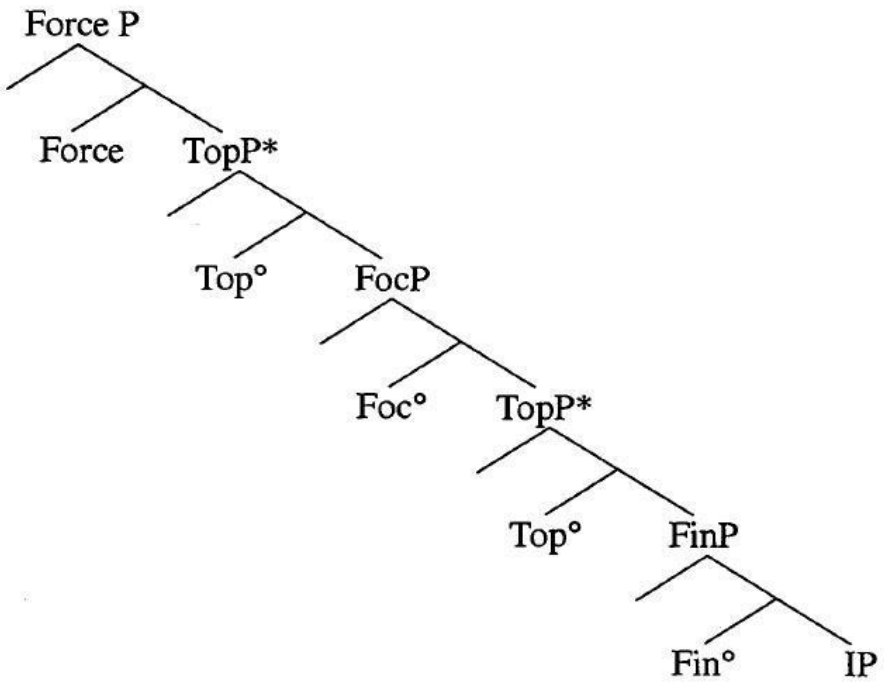
(30)

$AdvP_{speech\ act}$ (frankly,..)	$Mood_{speech\ act}$
$AdvP_{evaluative}$ (oddly,..)	$Mood_{evaluative}$
$AdvP_{evidential}$ (allegedly,..)	$Mood_{evidential}$
$AdvP_{epistemic}$ (probably,..)	$Mod_{epistemic}$
$AdvP_{past/future}$ (then,..)	$Tense_{past/future}$
$AdvP_{necessity}$ (necessarily,..)	$Mod_{necessity}$
$AdvP_{possibility}$ (possibly,..)	$Mod_{possibility}$
$AdvP_{habitual}$ (usually,..)	$Aspect_{habitual}$
$AdvP_{delayed}$ (finally,..)	$Aspect_{delayed}$
$AspectP_{predispositional}$ (tendentially,..)	$Aspect_{predispositional}$
$AdvP_{repetitive}$ (again,..)	$Aspect_{repetitive}$
$AdvP_{frequentative}$ (frequently,..)	$Aspect_{frequentative}$
$AdvP_{volition}$ (willingly,..)	$Mod_{volition}$
$AdvP_{celerative}$ (quickly,..)	$Aspect_{celerative}$
$AdvP_{anterior}$ (already)	$Tense_{anterior}$
$AdvP_{terminative}$ (no longer,..)	$Aspect_{terminative}$
$AdvP_{continuative}$ (still,..)	$Aspect_{continuative}$
$AdvP_{continuous}$ (always,..)	$Aspect_{continuous}$
$AdvP_{retrospective}$ (just,..)	$Aspect_{retrospective}$
$AspectP_{proximative}$ (soon,..)	$Aspect_{proximative}$
$AdvP_{durative/\ progressive}$ (briefly,..)	$Aspect_{durative/progressive}$
$AdvP_{prospective}$ (imminently,..)	$Aspect_{prospective}$
$AdvP_{obligation}$ (obligatorily,..)	$Mod_{obligation}$
$AdvP_{frustrative}$ (in vain,..)	$Aspect_{frustrative}$
$AdvP_{completive}$ (partially,..)	$Aspect_{completive}$
$AdvP_{manner}$ (well,..)	$Voice_{passive}$
Verb	Verb

A closer look at UG (Factor 1)

- Theoretical developments that have impacted on the formulation of parameters:

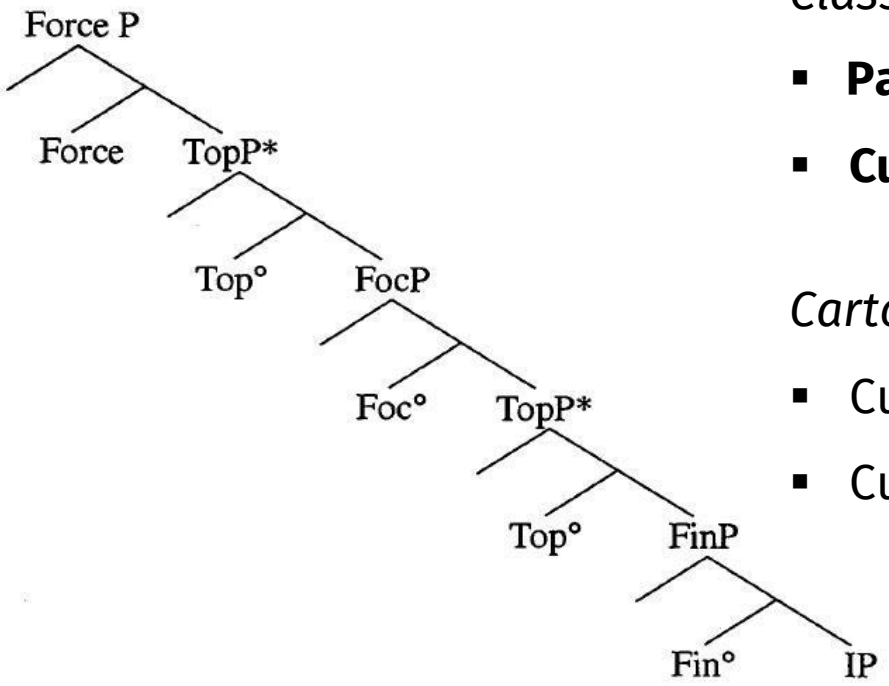
(29)



A closer look at UG (Factor 1)

- Theoretical developments that have impacted on the formulation of parameters:

(29)



Classic Verb Second (V2)

- **Parameter**: The finite verb **MOVES**/DOESN'T MOVE to C, the head of CP.
- **Cue**: $[_{CP}$ XP V-C $[_{TP}$ Subject …] (see (23a))

Cartographic Verb Second (V2)

- Cue for V2 in declaratives: $[_{TopP}$ XP V-Top … [Subj …
- Cue for V2 in *wh*-questions: $[_{FocP}$ *wh*XP V-Foc … [Subj …

(adapted from Lightfoot & Westergaard 2007: 410)

➢ microcues (Westergaard 2009), which potentially facilitate insight into the ways in which V2 systems change over time e.g. by becoming more/less "strict", or by restricting V2 in particular ways.

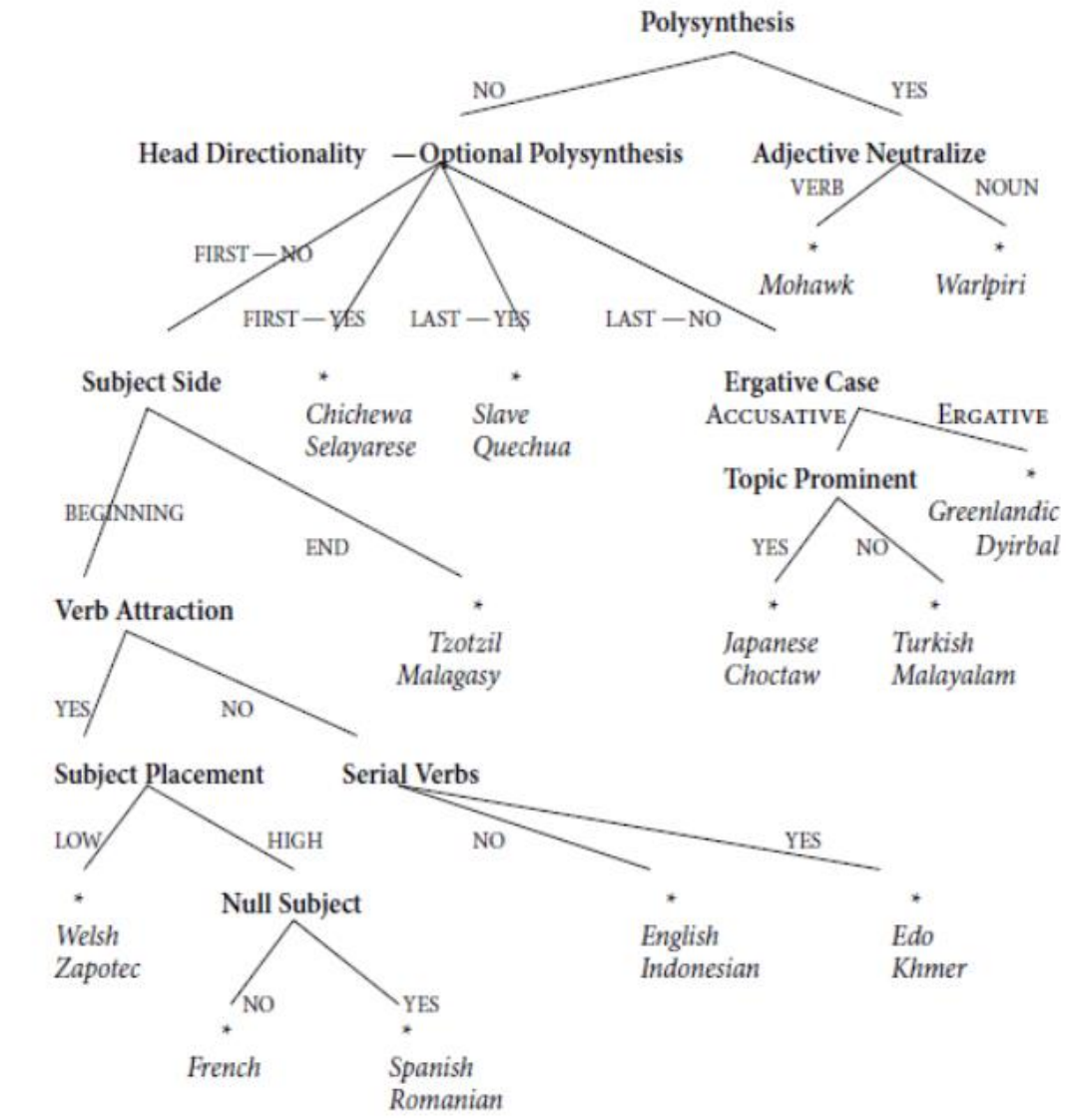
A closer look at UG (Factor 1)

- While there are various respects in which more fine-grained parameters can be a good thing (descriptive adequacy, among them), assuming the more fine-grained structures and associated parametric options to be part of UG poses other problems.
- Returning to the relationship between UG and the input:
  - Gibson & Wexler (1994)'s **Trigger Learning Algorithm** (TLA) (in)famously demonstrated that a trigger/cue/p-expression-based learner gets stuck on **local maxima** ('triggerless islands') even if there are only 3 simple parameters in play (OV/VO, V2 and Null Subjects).
  - To avoid this, parameters need to be **ordered** in relation to one another.
  - Recall (6):

'In a highly idealized picture of language acquisition, UG is taken to be a characterization of the child's pre-linguistic initial state. Experience – in part, a construct based on an internal state given or already attained – serves to fix the parameters of UG, providing a core grammar, guided perhaps by a structure of preferences and **implicational relations among the parameters** of the core theory.' (Chomsky 1981: 7)

A closer look at UG (Factor 1)

(31)



NB: These are ”basic” parameters

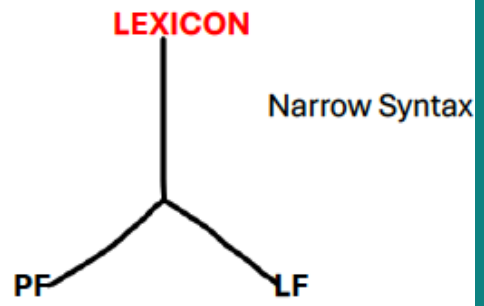
From Baker (2001) *The Atoms of Language*

A closer look at UG (Factor 1)

- There never was a model of the **interconnections** between parameters during the 2 Factors era.
- There was also never a **template** for what a parameter could look like (Gianollo, Guardiano & Longobardi 2008) and those postulated during the 1990s and early noughties often moved very far from the original ideal of 'parameterised principles', becoming very parochial in many cases.
- Crucially, P&P theory was never acquisitionally plausible: there was never a **learning theory** or an account of how the **Linking Problem** (see i.a. C.L. Baker 1979, Pinker 1984, Gervain & Mehler 2010, Fasanella & Fortun 2016) could be overcome.
- The model was also clearly not evolutionarily plausible, a topic that has become rather prominent in the Minimalist context (Berwick & Chomsky 2016).

A new approach to parameters in the Minimalist era

- The advent of Minimalism impacted the P&P perspective on parameters in that:



(i) it greatly simplified the architecture of the grammar

- no Deep vs Surface Structure; no modules

(ii) it became **feature**-centred:

- semantic and phonological features for the interfaces
- formal features (interpretable vs uninterpretable; later: valued/unvalued; or both) for NS
- functional categories (v, T, C, n, D …) = (structured) feature bundles
- movement was encoded via a diacritic (EPP-feature, edge feature, *, ^, etc.)

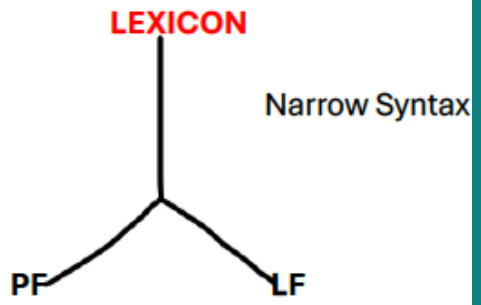
(iii) functional categories were meant to be limited in number

- Core Functional Categories (CFCs) – v, T, C in the clausal domain

➢ Many familiar parameters could be recoded in lexical terms.

A new approach to parameters in the Minimalist era

- Many familiar parameters could be recoded in lexical terms:



(32) Featural make-up of clausal head T:

(* signifies movement)

a. [V]	[D*]	--	English
b. [V*]	[D*]	--	French
c. [V*]	[D]	--	(Some varieties of) Arabic
d. [V]	[D]	--	Japanese

(33) **Borer-Chomsky Conjecture** (BCC; Baker 2008)

All parameters of variation are attributable to differences in the features of particular items (e.g. the functional heads) in the Lexicon.

- On the (classic) DM view, this means all non-roots.
- On the (current) Nanosyntactic view, roots would also be loci of (parametric) variation.

A new approach to parameters in the Minimalist era

- This new perspective on parameters can clearly become pretty fine-grained … and language (history)-specific … which isn't a natural fit for the classic view of parameters as innately pre-specified.

- It seems naturally suited to **microvariation** (Kayne 2005); but some properties don't look very "micro" - consider the OV typology in (26) again, for example.

- Baker (2008) also argues for the need to recognise **macroparameters**:

(34) A functional head F agrees with NP only if NP asymmetrically c-commands F.

(Yes: Niger Congo languages; No: Indo-European languages)

(35) A head F agrees with NP only if F values the Case feature of NP or vice versa.

(No: Niger Congo languages; Yes: Indo-European languages)

➢ Those both look like Agree-centred parameters, which would be a parameter-type that accords very well with the kind of templates for Minimalist parameters that have been proposed by Gianollo et al. (2008), Longobardi (2017) and Rizzi (2016).

A new approach to parameters in the Minimalist era

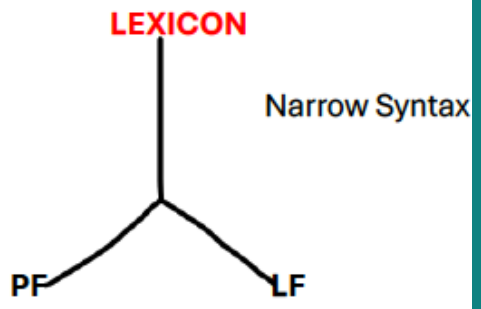
- Rizzi's (2016) Minimalist parameter types:

(36) a. **Merge Parameters**: Dictate how structures are built
e.g. establishing head-complement directionality or setting whether a language uses recursive combinations

b. **Move Parameters**: Govern how languages trigger and structure movement
e.g. A-movement , A'-movement, or whether a phrase must move to the edge of a phase for discourse criteria

c. **Spell-Out Parameters**: Determine which copies or structural positions are overtly pronounced vs. left unpronounced.

- These centre on the core Narrow Syntactic operations, Merge (= External Merge) and Move (= Internal Merge), and the on the PF component (Spellout)
  - deeper and more shallow parameters
  - All fundamentally rest on featural specifications, though, so this remains a fundamentally lexical approach to parameters



A new approach to parameters in the Minimalist era

- Longobardi's (2017) **parameter schemata** (building on Gianollo et al. 2008 and subsequent work and minimally resequenced)

(36) a. Is F, F a feature, grammaticalized ([F])?

b. Does [F] Agree with X, X a category?

c. Is [F] 'strong' (in the terminology of Chomsky (1995), i.e. does it trigger the movement of X)?

d. Does [F] move the minimal accessible category of type X (or is pied-piping possible)?

e. Is [F] spread onto X, X a category?

f. Does a functional category X have a phonological matrix Φ?

g. Are f1 and f2, the respective values of two grammaticalized features, associated on X, X a category?

h. Are f1 and f2, two feature values associated on X, optionally associated?

i. Does a functional feature (set) exist in the vocabulary as a bound/free morpheme?

➢ Again the objective is to formulate parameters referencing [F]s, the core operations and Spellout.

Feature-centred Minimalist and pathways of change

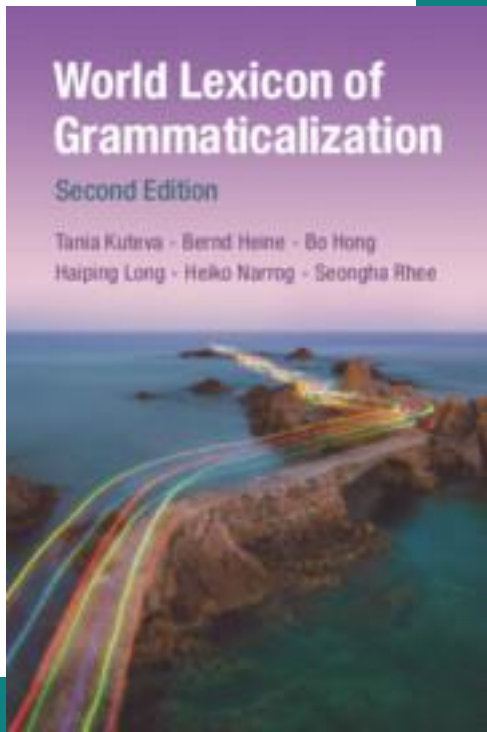
- Minimalism's feature-centred orientation facilitated the development of a rich generative approach to **grammaticalisation**.

- **Grammaticalisation**: creation of grammatical items from (originally) lexical ones (Meillet 1912, Hopper & Traugott 2003, Kouteva et al. 2019)

- It is crosslinguistically very robustly (universally?) attested (Kouteva et al. 2019)

(37) a. *have* ('possess') > perfect marker > past (V > Asp > T)
b. demonstrative > definite article (Adj > D)
c. *one* > indefinite article (Num > D)
d. *man* > indefinite pronoun (N > D) [*all vastly simplified*]

➢ **upwards** reanalysis (in the generative sense – see Roberts & Roussou 2003; recall Timberlake et al.)



Feature-centred Minimalist and pathways of change

- Minimalism's feature-centred orientation facilitated the development of a rich generative approach to **grammaticalisation**.
- Roberts & Roussou (2003) identify many cases instantiating **upwards** reanalysis:
  - (more) lexical items are merged at the bottom of the tree; (more) grammatical ones higher up.
  - In formal terms, this follows from an Economy condition, **F*merge over F*move**
    (see the Merge-over-Move idea that held in Minimalism in this period; more recently (Chomsky et al. 2023), it's Move-over-Merge …)
  - **Feature Economy**: Given two structural representations R and R' for a substring of input text, R is less marked (i.e., more economical and natural) than R' if and only if R contains fewer formal features than R'.
    - ➢ [F]-loss is desirable; and grammaticalisation is assumed to involve [F]-loss.

e.g. *have* the perfect/past marker has no thematic (argument structure) properties
a demonstrative-derived definite article lacks [deictic], etc.

[grammaticalisation also involves [F]-gain, though …]

Feature-centred Minimalist and pathways of change

- Minimalism's feature-centred orientation facilitated the development of a rich generative approach to **grammaticalisation**.
- Van Gelderen (2004) also identifies three Economy conditions shaping grammaticalisation, one featural (38), and the other two structural (39):

(38) **Feature Economy**: Minimise the semantic and interpretable features in a domain.

([S] > [iF] > [uF])

(39) a. The **Head Preference Principle** (HPP): Be a head, rather than a phrase.

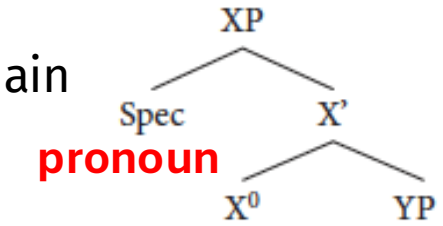
b. The **Late Merge Principle** (LMP): Merge as late as possible.

- e.g. of (39a): personal pronoun > agreement marker (agreement usually suffixal; so lower Spec and higher V, but prefixal agreement can also arise this way … though it seems to resist full reduction [prefixes look universally 'bigger' than suffixes –Julien 2002, Tokizaki & Kuwana 2013, Kayne 2017, etc.])

Feature-centred Minimalist and pathways of change

- Minimalism's feature-centred orientation facilitated the development of a rich generative approach to **grammaticalisation**.
- Van Gelderen (2004) also identifies three Economy conditions shaping grammaticalisation, one featural (38), and the other two structural (39):

(38) **Feature Economy**: Minimise the semantic and interpretable features in a domain ([S] > [iF] > [uF])



pronoun

agreement

(39) a. The **Head Preference Principle** (HPP): Be a head, rather than a phrase.

b. The **Late Merge Principle** (LMP): Merge as late as possible.

- e.g. of (39a): personal pronoun > agreement marker (agreement usually suffixal; so lower Spec and higher V, but prefixal agreement can also arise this way … though it seems to resist full reduction [prefixes look universally 'bigger' than suffixes –Julien 2002, Tokizaki & Kuwana 2013, Kayne 2017, etc.])

e

Feature-centred Minimalist and pathways of change

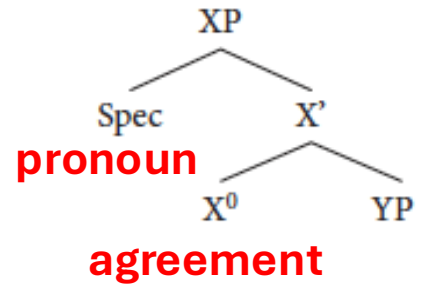
- Minimalism's feature-centred orientation facilitated the development of a rich generative approach to **grammaticalisation**.
- Van Gelderen (2004) also identifies three Economy conditions shaping grammaticalisation, one featural (38), and the other two structural (39):

(38) **Feature Economy**: Minimise the semantic and interpretable features in a domain.

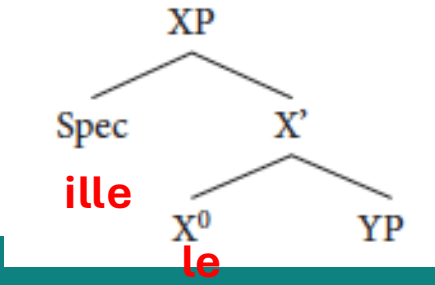
([S] > [iF] > [uF])

(39) a. The **Head Preference Principle** (HPP): Be a head, rather than a phrase.

b. The **Late Merge Principle** (LMP): Merge as late as possible.



- e.g. of (39a): personal pronoun > agreement marker (agreement usually suffixal; so lower Spec and higher V, but prefixal agreement can also arise this way … though it seems to resist full reduction [prefixes look universally 'bigger' than suffixes –Julien 2002, Tokizaki & Kuwana 2013, Kayne 2017, etc.])
- demonstrative > definite article; e.g. Latin *ille* as in *ille liber* 'that book' grammaticalised into French *le* as in *le soleil* 'the sun'.



- The LMP mostly replicates R&R's Merge-over-Move.

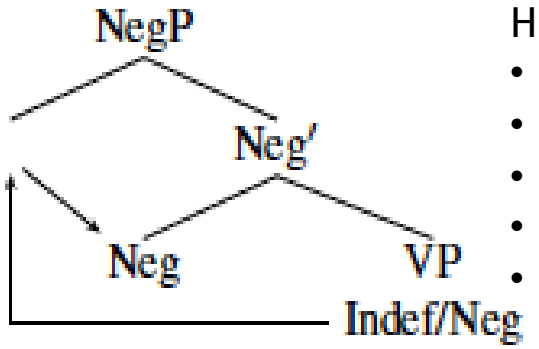
Feature-centred Minimalist and pathways of change

- Van Gelderen's Feature Economy also defines the formal path of **Cycles** (van Gelderen 2009, 2023).

Feature-centred Minimalist and pathways of change

- Van Gelderen's Feature Economy also defines the formal path of **Cycles** (van Gelderen 2009, 2023).

(40) The Negation Cycle (Jespersen 1917, after Alan Gardiner 1904)



Here:
- Stage I: Existing sentential negator in Neg (e.g. *ne*)
- Stage IIa: Optional reinforcement via [S]-bearing XP (e.g. *pas* = VP-internal Indef)
- Stage IIb: Optional reinforcer gains [ineg]-feature and moves to SpecNegP
- Stage IIIa: Obligatory reinforcement by [uneg]-feature-bearing XP, merged in SpecNegP
- Stage IV: Reanalysis of obligatory reinforcer as Neg

- The **microcycles** identified by van Gelderen (by 2023), alongside the negation cycle are:

  the Determiner Cycle, the Copula Cycle, the Tense and Aspect Cycle, the Mood Cycle, the Voice Cycle, the Interrogative Cycle, the Complementizer Cycle, and the Pragmatic Cycle (see also Mosegaard Hansen & Waltereit 2025 on cyclic developments in the discourse domain)

- Van Gelderen (2023) also identifies 2 **macrocycles**:

  the Agreement Cycle (head-marking) and the Case Cycle (dependent-marking)
  - these 'shift the typology of a language (van Gelderen 2023: 128) > Sapir (1921)'s 'drift'

Feature-centred Minimalist and pathways of change

- Returning to van Gelderen's Economy principles:
  - Initially, in van Gelderen (2004):

    '[they] are part of UG and help learners construct a grammar. They are similar to principles such as c-command, in that they remain active in the internalized grammar and therefore also aid speakers in constructing sentences. They are different from absolute principles such as c-command because prescriptive and innovative tendencies can counteract them.'

  - Van Gelderen (2009):
    - A link is made to other Minimalist Economy principles – Fewest Steps/Shortest Steps (≈Least Effort), Last Resort (Chomsky 1995 et seq.) – and also to Rizzi's (1990, 2001) Relativized Minimality, which is described as 'a natural principle of mental computation'
    - The Head Preference Principle is characterised as 'probably a more general cognitive principle, a third factor principle, "Analyse something as small as possible"'.

Concluding thoughts

Concluding thoughts

- The generative perspective on the formal make-up of grammar and how it is acquired has produced **significant new insights into long-established phenomena in diachronic syntax** during the 2 Factors era.
- Applying core generative theory diachronically **amplified** some of the **problems that also became evident in the synchronic domain** (e.g. the question surrounding the nature of cues/triggers and, more generally, the question of the structure of the input that child acquirers engage with).
- Additionally, there were indications from the start that 2 Factors, with just UG shaping I-languages, seemed insufficient: notions like the Transparency Principle and Degree-Zero Learnability don't naturally sit in UG.

➢ an early impetus within diachronic studies to acknowledge the need for Economy principles of some kind (contrast 'Some Notes on Economy of Derivation and Representation', which Chomsky started circulating in 1989, before further developing the ideas with Howard Lasnik, and popularising (!) them via *The Minimalist Program* (1995))

- ***Coming up (tomorrow):*** How the '3 Factors' era (Chomsky 2005; but, in practice, mostly post 2010) facilitates even more insight into both synchronic and diachronic questions (old and new), and accommodates both children and adults as drivers of change.