

The limits of syntactic variation: an emergentist generative perspective

Theresa Biberauer^{1,2}

University of Cambridge¹ & Stellenbosch University²

ReCoS Project (University of Cambridge)

The ReCoS project (2011-2016)

<http://www.mml.cam.ac.uk/dtal/research/recos/>



Rethinking Comparative Syntax

Emergentism within generative syntax

- Some familiar dichotomies:
 - **generativist** vs **constructivist**
 - **innate** vs **emergent**
 - categorial **universalism** vs categorial **particularism** (Haspelmath 2010)
- **BUT:** in the context of the ‘**3 factors**’ approach (Chomsky 2005)?
- **My objectives:**
 - to argue for an **emergentist approach to parametric variation** which has the capacity to offer new insights into **acquisition**, **crosslinguistic variation** (syntactic typology) and **change**.
 - to demonstrate that certain **pre-generative notions** also seem to have a central role to play in our understanding of language.

Generative Linguistics: updating the model

- “Traditional” UG model

(1) UG + PLD → adult grammar

- “rich” UG

(2) “FL [=UG; TB] specifies the features F that are available to fix each particular language L”
(Chomsky 2001: 10)

- BUT: “Current”/minimalist UG model:

(3) UG + PLD + **3rd factors** → adult grammar
(Chomsky 2005)

- impoverished UG → **How universal are minimalist features and categories?**

Generative Linguistics: updating the model

- The “new” factor:
 - (4) (a) principles of **data analysis** that might be used in language acquisition and other domains;
 - (b) principles of **structural architecture and developmental constraints** that enter into canalization, organic form, and action over a wide range, including principles of **efficient computation**, which would be expected to be of particular significance for computational systems such as language. (Chomsky 2005:6)
- Here: 3rd factors = generally applicable **learning biases**
- (5) **Make maximal use of minimal means**

Generative Linguistics: updating the model

- Linguistic manifestations of this general learning bias:
 - (6) **Feature Economy (FE)**: postulate as few features as possible to account for the input (=intake) [generalised from Roberts & Roussou 2003]
 - (7) **Input Generalisation (IG)**: maximise available features [generalised from Roberts 2007]
 - ➔ **minimax search/optimisation algorithm**
 - (8) a. Factor 1 = UG: basic operations, Merge and Agree, plus a formal feature template [iF]/[uF] and a very small subset of [F]s not derivable from the input
 - b. Factor 2 = PLD (intake), particularly evidence of movement, doubling, systematic silence and multifunctionality
 - c. Factor 3 = FE and IG

The plan for today's talk

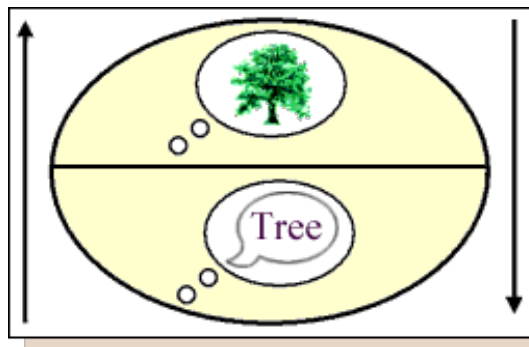
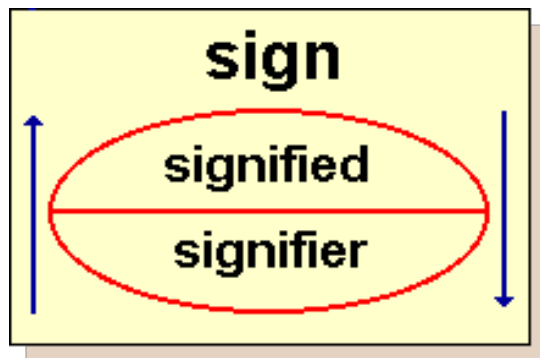
- **Part I:** a model of how 3 factors may shape language variation
- **Part II:** predictions regarding the shape of variation and how this parallels what we see outside of language
- **Conclusions:** typological, acquisition and diachronic perspectives

PART I:

A model of how 3 factors may
shape language variation

The role of UG (Factor 1)

- Acquirers learn arbitrary sound-meaning mappings
- Saussurean **arbitrariness**:
 - words are **signs**
 - the **signifier** (*signifiant*) - the **form** of the sign
 - the '**signified**' (*signifié*) - its **meaning**

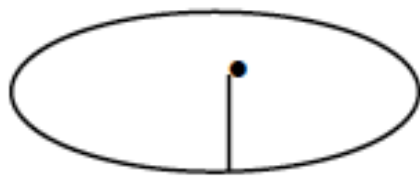


Ferdinand de Saussure
1857-1913)

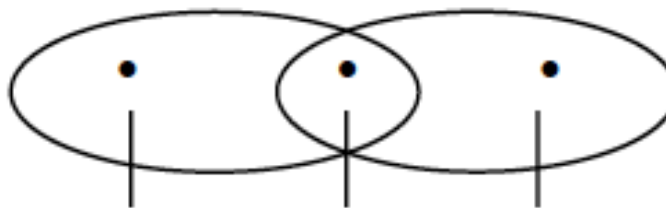
The role of UG (Factor 1)

- But lexically specified sound-meaning mappings are not enough (the essence of human languages?)
- **Proposal:** a **UG-given [uF]/[iF] feature template** guides acquisition of **syntactic features and, consequently, categories**

Phonological features Formal features Semantic features



[P]



[uF]

[iF]

[S]

(cf. Zeijlstra 2008)

The mechanism: non-[P]- and [S]-related **regularities** are systematically encoded in [iF]/[uF] terms in all languages → **universality**

The role of UG (Factor 1)

... but the features ([F]s) are **emergent**; they do not derive from a pre-specified UG inventory (*pace* Chomsky 2001).

- even before 3 factors: ‘universal’ was interpreted in ways that produced very different universal categories (Ramchand & Svenonius 2014):
 - **Rich Functional Hierarchies** (RFH): cartography, nanosyntax (Rizzi 1997, Cinque 1999, etc.; the Tromsø project)
 - **Minimal Role for UG** (MUG): roots, categorisers as in Distributed Morphology (Marantz 1997 *et seq.*) and some functional categories (phase-heads)
- Recent work, notably that of Wiltschko (2014) and colleagues, highlights the problematic nature of a UG-given functional sequence (Cinque 2013)

Features and Categories without UG?

- No (complete) UG-given inventory of formal features ([F]s)

BUT: Where do [F]s come from?

- [F]s and the categories they define result from the interaction of the 3 factors
- (9) a. Factor 1 = UG: basic operations, Merge and Agree, plus a formal feature template [iF]/[uF]
- b. Factor 2 = PLD (intake), particularly evidence of movement, doubling, systematic silence and multifunctionality
- c. Factor 3 = FE and IG
- Crucially, **the PLD will “skew” acquirers’ postulation of [F]s**, depending on what is salient in a given language → We should not expect (many) universally formally identical categories (cf. also Wiltschko 2014)

The role of the PLD (Factor 2)

- **Key cues** = systematic departures from Saussurean arbitrariness
e.g. movement, doubling, systematic silence & multifunctionality
(cf. also Zeijlstra 2008 and much recent work by Martina Wiltschko)

- The logic:
 - **doubling**, e.g. agreement = 2/multiple forms, 1 meaning

(10) la belle fille [French]
the.FM.SG pretty.FM.SG girl.FM.SG
'the pretty girl'

(11) Hulle is **nie** laat **nie** [Afrikaans]
 they is NEG late NEG
 ‘They are not late’

The role of the PLD (Factor 2)

- **systematic silence**, e.g. null exponence, ellipsis (**no form with meaning**)

(12) a. Chomsky \emptyset wrote a new paper (didn't he?)
b. Did Chomsky write a new paper?

(13) a. ona vrač [Russian]
 she doctor = 'She is a doctor'
b. on **byl** učenik-om
 he be.M.PST pupil -INSTR = 'He was a pupil'

(14) a. ő diák [Hungarian]
 3SG.M pupil = 'He is a pupil.'
b. én tanár **vagyok**
 1SG teacher be.1SG.PRES = 'I am a teacher'

CONTRAST

The role of the PLD (Factor 2)

- **multifunctionality** – 1 form, multiple meanings (Wiltschko 2014, Duffield 2013, 2014)

(15) a. Ong Quang **đu'ợ'c** mua cái nhà
Prn Quang CAN buy CL house
'Quang is allowed to buy the house'

Deontic

b. Ông Quang mua **đu'ợ'c** cái nhà
Prn Quang buy CAN CL house
'Quang bought a house'

Accomplishment

c. Ông Quang mua cái nhà **đu'ợ'c**
Prn Quang buy CL house CAN
'Quang can buy a house'

Epistemic/Abilitative

More **CONTRAST**

Vietnamese

The role of the PLD (Factor 2)

- **movement** – assuming Chomsky’s (2000) notion of **duality of semantics** (thematic + discourse/scopal meaning), movement results in “extra” meaning

- (16) a. [Never in my life] **did** I expect that to happen!
b. [With no job] **would** she be happy [Modern English]

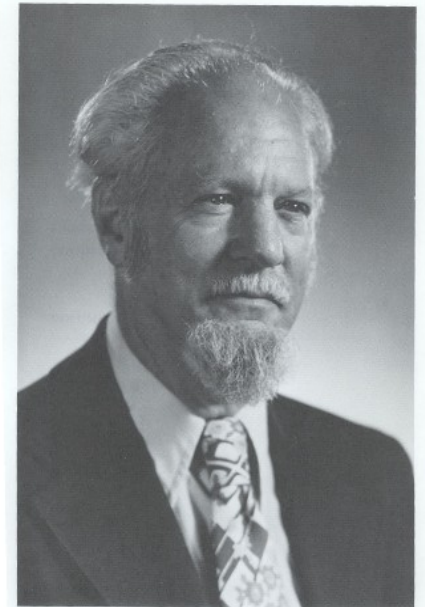
- (17) we may **non answer** 3eue to pys pynge þat he hath seide
we may no answer give to this thing that he has said
‘We can give no answer to this thing that he has said’

(*Barlam and Iosaphat* (EETS O.S. 290), 5597; from Mackenzie & van der Wurff 2012: 850)

- Also relevant here: the ‘higher-level’ **duality of patterning** deriving from “neutral” vs “marked” word orders

The role of the PLD (Factor 2)

- “standard” **duality of patterning**:
 - meaningless phonemes: /t/, /r/, /i:/, etc.
 - meaningful phoneme-combinations: *tree*, etc.
- “higher-level” **duality of patterning** (cf. also Fortuny 2010)
 - meaningless first headedness choice: “basic”
OV vs VO
 - meaningless obligatory filling choices: V spellout
position, Spec-TP, Spec-CP, etc.
 - meaningful optional movements relative to the fixed
higher-level conventions: e.g. T-to-C in English, the
nature of the XP that raises to Spec-CP, etc.
- Having both levels of duality of patterning = **maximising the contribution of both the Lexical Items and (External and Internal) Merge (= minimal means)**



CHARLES F. HOCKETT

Charles Hockett
(1916-2000)

The role of the PLD (Factor 2)

- **The driving intuition here:** [F]s are postulated if they can be seen to regulate some form of **systematic contrast**, which cannot be explained by appealing only to semantic or phonological considerations (a higher level of Saussurean arbitrariness)
 - The morphosyntactic and morphosemantic contrasts vary by language; hence the language-specific “content” of what it means to “be” categories of different types, and also what features are grammaticalised (i.e. [F]s) is expected to vary (cf. also i.a. Ritter & Wiltschko 2009, 2010, 2014, Wiltschko 2014, and Chung 2012 on this)
- categories and features are **emergent**
- **BUT:** doesn't this just predict rampant and unconstrained variation?

The role of Factor 3

- The general (non-language-specific) learning bias:

(18) **Maximise minimal means**

→ 2 linguistic manifestations:

(19) **Feature Economy (FE)**: postulate as few features as possible to account for the input (=intake)

(20) **Input Generalisation (IG)**: maximise use of postulated features

The role of Factor 3

- Linguistic evidence of FE at work

- “recycling” effects

(21) a. John-han^{they} Mary-ka mwusewe. [Korean]

John-DAT Mary-NOM be.afraid

b. John-han^{they}-**ka** Mary-ka mwusewe.

John- DAT- **NOM** Mary-NOM be.afraid

'JOHN is afraid of Mary.'

(via Pesetksy 2014)

→ “case-stacking” (cf. Richards 2013, Levin 2013, Pesetksy 2014)

The role of Factor 3

- More **linguistic evidence** of FE in play (“recycling”):
 - multi-domain/“extreme”: use of agreement (e.g. Chamorro, Archi)
 - “monstrous” agreement in Tamil and other languages (Sundaresan 2012)
 - “doubling” pronouns (Leffel, Simik & Wierzba 2013)
 - “double life” co-ordination markers in Japanese, Malayalam, etc. (Jayaseelan 2015, Mitrović 2015)
 - “multi-life” focus/question/polarity, etc. particles crosslinguistically
 - grammaticalization phenomena generally

The role of Factor 3

- More **general evidence** of FE (Mobbs 2015):
 - learners seem to characterise ('parse') their input using the minimum set of postulates (Gallistel & Gibbon 2000, 2002, Gallistel 2002, 2003)
 - diachronically, "outlier" elements frequently lost (Biberauer & Roberts 2014, tomorrow)

The role of Factor 3

- Linguistic evidence of IG at work

Acquisition of English *wh*-movement (Thornton 1995)

- (22) a. **Who** do you think **who**'s under there?
b. **What** do you think **what** Cookie Monster eats?
c. **How** do you think **how** Superman fixed the car?

→ *wh*-copying

- (23) a. **Stage I (initial hypothesis)**: Spec-head agreement = obligatory (IRISH)
b. **Stage II**: Spec-head agreement = obligatory for subjects (FRENCH)
c. **Stage III**: Spec-head agreement = obligatorily covert for subjects

The role of Factor 3

- More **linguistic evidence** of IG in play:
 - root infinitives (over-generalisation of most common form in base position)
 - “shadow” phenomena (Demuth 1994, 2003; Lleó 1998, Lleó & Demuth 1999, van Kampen 2004)
 - “harmony”/“contiguity” phenomena:
 - FOFC (Biberauer, Holmberg & Roberts 2014, Sheehan 2013)
 - the behaviour of mixed extended projections (Panagiotidis 2014)
 - (non)agreement “cut-off” effects (Pesetsky 2014, Ouwayda 2014)
 - Animacy and Case hierarchy (Caha 2009) effects

The role of Factor 3

- More **general linguistic evidence** of IG
 - lexical overgeneralisation: *yesterday* = “any time in the past” vs *Saturday* = “any time in the future”
 - morphological overgeneralisation: *went* > *goed* > *went*
 - “specialised” use of C and V, and stress in acquiring lexicon and morphosyntactic regularities (Mehler and colleagues)
 - experimental evidence, e.g. Hudson Kam & Newport (2005): under experimental conditions, children learning SillySpeak regularize variable input in various ways, while adults do not;
➔ “children learn unpredictable variation differently than adults. They have a stronger tendency to impose **systematicity** on inconsistent input ...” (Hudson Kam & Newport 2005:184)

The role of Factor 3

- The **types of regularization** that children impose on the input:
 - (24) a. **maximization**: use the variable form all the time
 - b. **minimization**: use the variable form none of the time
 - c. **linguistically governed selection**: use the variable form in a grammatically defined subset of contexts (e.g. only with transitive Vs)
- Children do similar things when exposed to complex variability in probability learning experiments (e.g. a Left, Middle and Right strategy where there are three lights that variably flash)

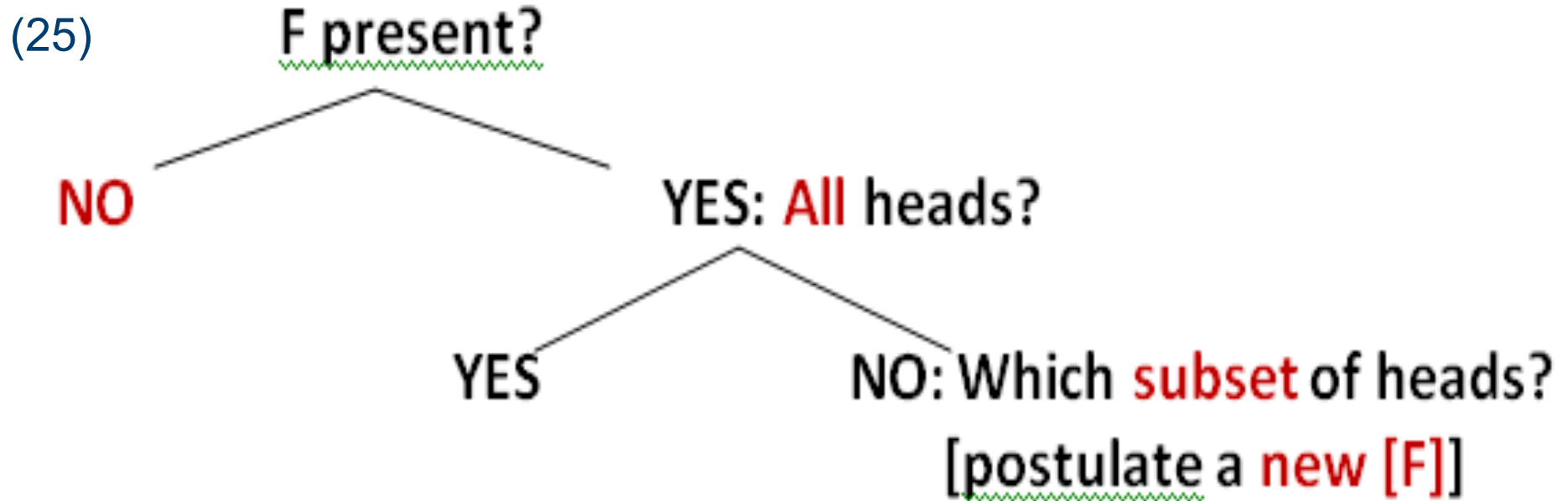
Part II

Predictions: the shape of **variation** and how this parallels what we see outside of language

The interaction of the 3 factors and variation

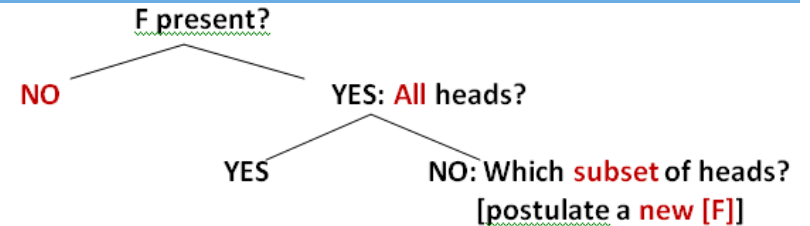
Constraints on features and categories

- Recall: together, FE and IG form a **minimax search/optimisation algorithm** (FE: minimise features; IG: maximise detected features).
- They also naturally result in a **learning pattern/path** (hierarchy) with the following general “shape”:



The interaction of the 3 factors and variation

- **NO > ALL > SOME** acquisition sequence
- **NO** = default as the acquirer doesn't receive (systematic) input pointing to need for F **FE & IG respected**
- **ALL**: plausibly follows from the acquirer's initial "ignorance" (Biberauer 2011, Branigan 2012) **IG respected & FE minimally violated ("Make maximal use of minimal means" – cf. Biberauer 2011)**
- **SOME**: both IG & FE violated, but FE will dictate that the violation should be minimal, picking up on the nature of the input cues
- "Recovery" from **superset traps** (Berwick 1985) possible as the acquirer must postulate appropriately specified heads to capture the data, with the [uF]/[iF] template in combination with IG and FE gradually reducing the acquirer's "ignorance"



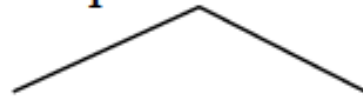
A hypothetical head-directionality parameter hierarchy

(26)

Is head-final present?



No: **all head-initial** Yes: present on all heads?



Yes: **all head-final**

No: present on all [+/-V] heads?



Yes: No: present on a subset of [+/-V] heads? ...

head-final in the clause/nominal

NO>ALL>SOME outside syntax

- The **Successive Division Algorithm** (Dresher 2009:16):

- (27) a. Begin with *no* feature specifications: assume all sounds are allophones of a single undifferentiated phoneme.
- b. If the set is found to consist of more than one contrasting member, select a feature and divide the set into as many subsets as the feature allows for.
- c. Repeat step (b) in each subset: keep dividing up the inventory into sets, applying successive features in turn, until every set has only one member.

NO>ALL>SOME outside syntax

- Jaspers' (2013) **Concept Formation Constraint**:

- (28) 'a set of four natural operators is generated by making subtractions from a fixed domain space of values via a series of **two successive binary divisions**. There is an initial exhaustive division between the contradictories NOR and OR ... within the remaining non-NOR space of values, we can either carve out the subset AND, leaving inclusive OR as superset space ... or we can divide the inclusive OR space exclusively into AND and exclusive OR'.
- (29) Predicate calculus oppositions: $[[_{\text{Some}2} \text{All}; \text{Some}_1] \text{None}]$
- (30) Propositional calculus oppositions: $[[_{\text{Or}2} \text{And}; \text{Or}_1] \text{Nor}]$

NO>ALL>SOME outside syntax

(31)

a. *Domain*

1	1
1	0
0	1
0	0

All truth-value pairs

b. *Step 1*

1	1
1	0
0	1
0	0

or

nor

Contradiction
Something is true
vs.
Nothing is true

c. *Step 2*

1	1
1	0
0	1
0	0

and

incl. or

Implication
Something is true
vs.
Everything is true

d. *Step 2'*

1	1
1	0
0	1
0	0

and

excl. or

Something but not
everything is true
vs.
Everything is true

Predictions: typological predictions

- There are **constraints** on the features and categories that acquirers will postulate, i.e. **limits on featural and categorial variation**
- Assumption: grammatical structure (and possibly also phonological and morphological structure – cf. Nevins 2010, Halle & Marantz 1997) is created via the operations **Merge** and **Agree**

(32) Impossible (full/non-pidgin, etc.) systems

- a. [F]-less systems
- b. systems containing [F]s not independently responsible for some instance of doubling, movement, “silence” or multifunctionality
- c. systems lacking a basic distinction between “spine” elements (e.g. verbal heads making up the clausal spine) and “satellite” elements (e.g. subjects, objects, certain adverbials)

Predictions: typological predictions

- In a model where **salient [P] and [S] input** counts heavily in determining how children get their [F] systems off the ground, we expect there to be crosslinguistic variation regarding categorial make-up – **even** where lexical categories are concerned: [V]/[N] are not given
- That ϕ -features could be [uF] on verbs in one system and [iF] on verbs in another is, for example, readily possible on the present model.
- Similarly, features that are not typically thought of as **category-defining** may emerge as such on this model.

E.g. in languages where V and N differ in initiality/finality (disharmonic languages), headedness may be an [F] distinguishing verbal vs nominal categories (Biberauer 2013, 2014, Biberauer, Holmberg & Roberts 2014)

Predictions: typological predictions

- PLD: **ordering information** = salient → **basic word-order properties** fixed early (Wexler's Very Early Parameter Setting)
- **Proposal:** headedness (finality/initiality) = a **category-defining** property. i.e. in V-final languages, [^] is *part of* what it means to “be a (lexical) verb/verbal”

[This has important implications for our understanding of FOFC as defined in Biberauer, Holmberg & Roberts 2014]

- In contrast, **substantive formal features**, which (mostly) have to be acquired on the basis of morphosyntactic cues (movement, agreement, systematic “silence”, multifunctionality) are typically **category-refining**, facilitating sub-distinctions between categories.
 - e.g. the presence of ϕ -features may distinguish category T from the category of lexical verbs (T: [V[^], ϕ] vs V: [V[^]]) (and T may additionally lack V's [^] (T: [V, ϕ] vs V: [V[^]], as in Vata, Nupe, etc.)

Typological predictions

- A general expectation: languages displaying **different sized versions of the same phenomenon**
- Also, if we compare languages, the SOME choices can take different forms:
 - (i) they may become more successively more constrained, each option being a **subset** of the previous one (e.g. ever more restricted domains for head-finality); or
 - (ii) they may not be in a featural subset relationship (e.g. the choice as to whether [N] or [V] heads are final, once it's been established that there is a headedness difference in the clausal and nominal domains, or the extension vs restriction option in the alignment parameters discussed in Sheehan (2015))
- (ii)-type SOME options are **at the same level of a parametric hierarchy** as they are **not successively considered by acquirers**; from a typological perspective, they are equivalent, i.e. we may be able to develop some sense of what **typological equivalence classes** look like.

Diachronic predictions

Expectations:

- pressure to postulate **simpler systems**, i.e. extend a generalization over the widest possible domain – a common outcome where PLD present complexity of some kind (e.g. variability, but not only this; cf. the expansion of V2 in Afrikaans; Biberauer 2014, 2015)
- but also the converse: situations under which the presence of an erstwhile [F] becomes less transparent, leading to “domain restriction” on account of the harnessing/introduction of a further [F] to account for a more limited instantiation of a previously (more) productive process (Biberauer & Roberts 2009 on the loss of OV order, and Biberauer & Roberts 2014 on the gradual reduction in Conditional Inversion in English).

Diachronic predictions

- Lesser transparency may ultimately lead to the loss of an [F] and thus simplification (which may result in an upward “leap” within a given parametric hierarchy; see again Biberauer & Roberts 2014 on the complete loss of Conditional Inversion and, indeed, Inversion in certain “New Englishes”)

i.e. in diachronic terms, we don’t expect unidirectional or even “contiguous” movement along a parametric hierarchy

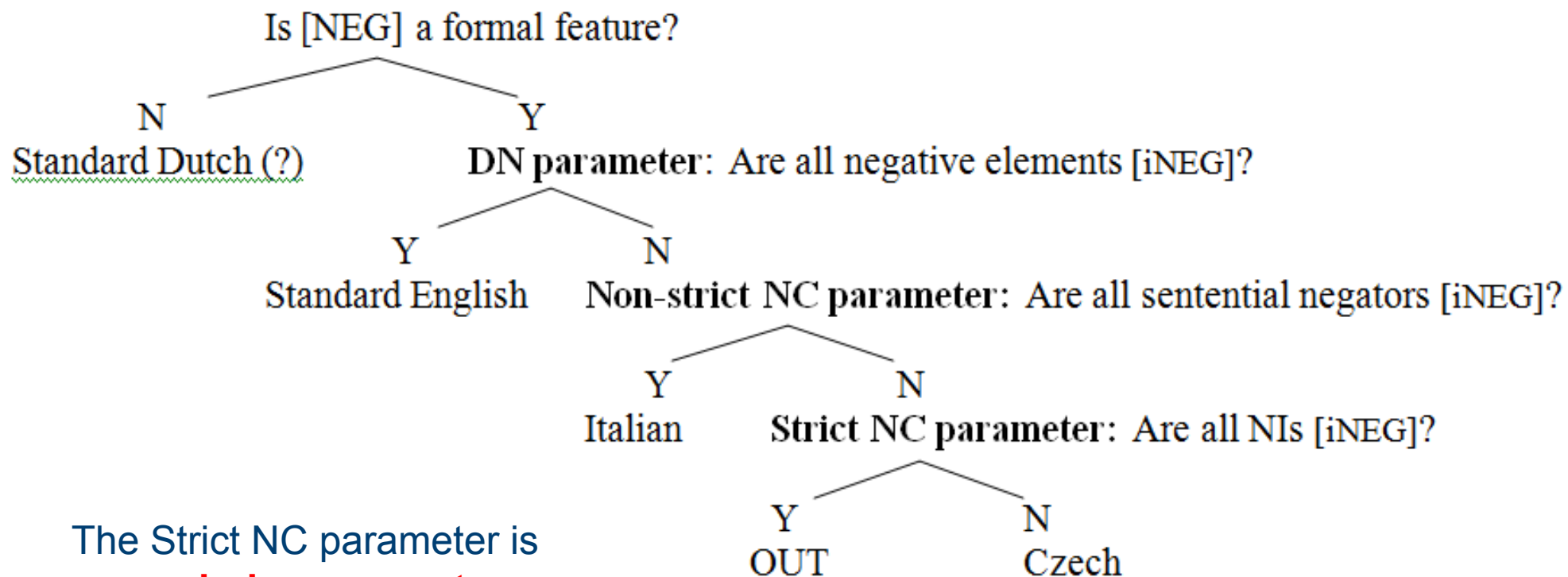
- We expect **contact situations** to be very interesting for formal reasons
- And also those in which strong normative influences are (initially) absent

Conclusions

- Minimalistically oriented generativists do not have to subscribe to an **aprioristic, universalist** position on syntactic categories (or features).
- Taking into account (i) “traditional” conceptual and empirical problems with strong universalist positions (e.g. the Linking Problem, crosslinguistic diversity), (ii) independent arguments for emergent features in phonology (e.g. Mielke 2008, Dresher 2009), and (iii) the new possibilities opened up by the 3 factors framework (Chomsky 2005), adopting an **emergentist approach to syntactic features** seems extremely well motivated.
- This type of approach predicts greater crosslinguistic variation than before, BUT it also predicts that this variation – as manifested in **synchronic typology**, **diachronic change** and **acquisition** – will be **restricted** and **take on a very specific character**.
- The approach also seems to have consequences for the **formal effects of language contact** (cf. Biberauer & Roberts 2012, 2014, Biberauer 2014, 2015)

A parametric hierarchy for negation

(33) The feature decision tree reformulated in explicitly parametric terms



The Strict NC parameter is
a **no-choice parameter**