# Rethinking Comparative Syntax
# Part One: Introduction

Ian Roberts

University of Cambridge

igr20@cam.ac.uk

# Part One: Rethinking Comparative Syntax

- ☐ The Principles and Parameters approach to comparative syntax has suffered some criticism in recent years;

- ☐ However, proposals to abandon it (Newmeyer 2005, Boeckx 2011, etc) don't really offer viable alternatives

- ☐ Giving up on accounting for variation is a bad idea → hence the Rethink

# The "classical" view (Chomsky 1981)

Universal Grammar (UG) contains:

- ☐ fixed invariant **principles** (e.g. a head X is sister to its complement YP in X');

- ☐ **parameters** of variation (e.g. X precedes/follows YP in X').

NB *both* principles and parameters are part of the innate linguistic endowment.

# The explanatory value of P&P

- ☐ Simple general principles, not needing to be learnt;
- ☐ Simple parts of Primary Linguistic Data (PLD) trigger abstract parameter values;
- ☐ A solution to the poverty-of-the-stimulus problem
- → **Explanatory adequacy (Chomsky 1964).**

# The scope of the P&P approach

- ☐ language typology: parameters make predictions about (possible) language types, e.g. head-initial X>YP vs head-final YP > X;

- ☐ L1 acquisition is parameter-setting;

- ☐ language change is parameter change.

# Rethinking Parameters

- ☐ The Minimalist approach doesn't really have principles that can be parameterised;

- ☐ Question of "granularity" of variation (macro- and microparameters), see below;

- ☐ Reducing the innate endowment (Occam's razor plus general plausibility).

# **Micro- vs. macroparameters**

(1) The Head Parameter:

a. X' → {X, YP}
b. X precedes/follows YP

(2) Is John coming? He said he might *US(do).
(Non-finite *do*-insertion in US vs. UK English: Kayne (2005:7)).

# In favour of microparameters

(3)    The "Borer-Chomsky conjecture" (BCC):
- ☐ All parameters of variation are attributable to differences in the features of particular items (e.g. the functional heads) in the Lexicon. (Baker (2008a:3; 2008b: 156)

(4)    *Formal* features of functional heads:
- ☐ Case, person/number/gender (φ-features), categorial features; movement-triggering feature(s)

# I.   A strong limit on what can vary:

(5)   Impossible parameters (which could have been countenanced under a GB approach, *mutatis mutandis*):
a.     "Arity" of Merge ($\rightarrow$ non-configurationality);
b.     existence of movement (given Move= Internal Merge);
c.     mode/level of lexical insertion.

# II. An argument based on acquisition:

(6) Associating parameter values with lexical entries reduces them to the one part of a language which clearly must be learned anyway: the lexicon.

(Borer (1984:29))

☐ NB this is true even if, perhaps especially if,  the UG (first-factor) component in acquisition is radically reduced (cf. Biberauer 2013).

# III. A restriction on the form of parameters

(7) For some formal feature F, P = ±F.

(8) a. Non-finite T is [±φ]: Euro. Port. inflected infinitives vs. English, etc.

b. N is ±Num: Chinese vs. English, etc.

c. T is ±EPP: determines position of subject (Welsh vs. English).

☐

# **Underspecification**

□   This simplicity of formulation in turn makes possible a statement of parametric variation at the UG level which relies on the logic of underspecification:

(9)a.   For some formal feature f, -f is the default value of P.

   b.  P has +f when triggered (i.e. under specified conditions), -f elsewhere.

   c.   +f is the marked value of P.

**AND the set F of possible formal features ($f_1$.. $f_n$) is not necessarily given in advance (see again Biberauer 2013)**

# **Microparametric approaches**

- ☐ Have the advantages just listed
- ☐ Tend to be descriptively more adequate
- ☐ But postulate too many parameters, undermining the supposed explanatory value of the approach.

# Macroparameters: Baker (2008)

- ☐ [on the microparametric view] "there should be many mixed languages of different kinds, and relatively few pure languages of one kind or the other" (10).

- ☐ On the other hand, the macroparametric view predicts, falsely, rigid division of all languages into clear types (OV vs VO etc): every category in every language should pattern in one way or the other (see (1)).

# Combining the macroparameters and microparameters

□ we expect to find a bimodal distribution: languages should tend to cluster around one type or another, with a certain amount of noise and a few outliers from either one of the principal patterns:

OV / Postpositions 472

OV / Prepositions 14

VO / Postpositions 41

VO / Prepositions 427

(Dryer 2011:maps 83A and 85A)

# The Romance languages

- ❑ home of the microparameter?
- ❑ A well-documented group of fairly closely related languages
- ❑ Nothing special (synchronically)
- ❑ Just better-studied, esp. from this perspective, than most families

# a. Subject clitics

i.  a dɔrmə, tə dɔrmə, i/al dɔrmə, a durmiŋ, durmitə, i/al dɔrmənə (Carrara, N. Italy)

ii.  dɔrmi, ta dɔrmat, al/la dɔrma, dɔrmum, dɔrmuf, dɔrmaŋ          (Como, N. Italy)

iii.  je dors, tu dors,  il/elle dort,  nous dormons, vous dormez, ils/elles dorment  (French)

iv.  .. and true NSLs have none at all (Standard Italian, Spanish, Portuguese)

# b. Negation:

i.    Non dormo.          (Italian)

ii.   Je ne dors pas.     (French)

iii.  i drøma mia   (Trecate, N. Italy)

     "I don't sleep."

# c. Enclisis:

i.    Le faire demain ...    (French)

ii.   Farlo domani..    (Standard Italian)
      "To do it tomorrow"

iii.  A Maria viu-o.    (European Portuguese)
      "Mary saw him."

iv.   Nenhum aluno o viu.    "
      "no student saw him"

# d.   past-participle agreement :

i. La manzana ha sido comida.   (Spanish)
   "The apple has been eaten-Agr"

ii. Je l'ai peinte (la table)            (French)
    « I've painted-Agr it, the table »

iii.    Giuwanne a pittite ddu mure
    (Abruzzese, C. Italy)
    'John has painted-Agr two walls'

# e.   the nature and choice of aspectual auxiliaries:

i.    He llegado.          (Spanish)
      Have arrived

ii.   Sono arrivato.       (Italian)
      "Am" arrived.

iii.  Tu si fatte na torta.    (Abruzzese)
      You are made a cake.

      Esse a fatte na torta.
      He has made a cake

# BUT all Romance languages:

(18)    a.        are SVO;

             b.        are Prepositional;

             c.        are non-ergative (although some "split-ergativity" is attested in auxiliary selection in C/S Italy);

             d.        are morphologically fusional;

             e.        fail to show the Chinese value of Chierchia's (1998) Nominal Mapping Parameter (*Ho visto gatto "I saw cat" referring to a single instance of a cat);

             f.        have definite and indefinite articles;

             g.        have moderately rich agreement systems;

             h.        have complement clitics  (except a small number of Rhaeto-Romansch varieties; see Benincà & Poletto 2005);

☐        i.        fail to show a full morphological case system.

# **Questions for classical parameter theory**

☐ What do we say about proclisis-enclisis parameters in systems with no clitics (e.g. Germanic, Chinese, Japanese, Korean ..)?

☐ What do we say about participle-agreement parameters in systems without participles or agreement?

☐ If such variation is not conditioned by parameters, what is it?

# A taxonomy of parameters

For a given value $v_i$ of a parametrically variant feature F:

a. **Macroparameters**: all heads of the relevant type share $v_i$;

b. **Mesoparameters**: all heads of a given natural class, e.g. [+V], share $v_i$;

c. **Microparameters**: a small, lexically definable subclass of functional heads (e.g. modal auxiliaries, pronouns) shows $v_i$;

d. **Nanoparameters**: one or more individual lexical items is/are specified for $v_i$.

# Macroparameters

- ❑ Gross typological properties which tend to hold at the genus level or higher (because they're strongly conserved), e.g. for Romance:

- ❑ SVO+Prepositions (general head-initiality);

- ❑ Accusative alignment;

- ❑ morphological fusion;

- ❑ definite and indefinite articles and (perhaps concomitant) failure to show the Chinese value of Chierchia's (1998) Nominal Mapping Parameter.

# Mesoparameters

- ☐ Moderately rich agreement and (concomitant?) consistent null subjects (except North-Western varieties which have subject clitics)

- ☐ Pronominal object clitics (except basolectal Brazilian Portuguese)

# **Microparameters**

- ☐ Subject clitics (NW)
- ☐ Negation (Jespersen's cycle)
- ☐ Proclisis-enclisis alternations (esp. EurPrt)
- ☐ Past-participle agreement (esp. S/C Italy)
- ☐ Aux selection (esp. S/C Italy)

# Nanoparameters

□ French *combien* violates LBC:

Combien as-tu lu [ (combien) de livres ] ?
How-many have-you read        of books?


□ Romanian 3sg.f. clitic –*o* is enclitic to lexical
  V in compound tenses:
  Am văzut – o.
 have-I seen 3ACC FEM
  'I saw her'.

# The central idea

☐ Macroparameters are the result of aggregates of microparameters acting together, effectively as a single parameter.

☐ Mesoparameters are slightly smaller aggregates

☐ Nanoparameters are lexically-governed exceptions to prevailing values in the system

☐ (the notion of parameter combined with the notion of set – Kayne).

# Parameter hierarchies: word order

☐ The "head parameter":

a.  Are movement-triggering features present in the system?   (Y/N)
   N: rigidly, harmonically head-initial language (Welsh)

b.  Y: If so, are these features obligatory on all heads?
      Y: rigidly, harmonically head-final language (Japanese)
c.  N: If not, are these features present on some subset of heads e.g. [+V]?

# Parameters and the three factors

☐ F1: where UG doesn't mind (underspecification);

☐ F2: trigger experience;

☐ F3: general strategies of L1 acquisition based on computational conservatism.

Parameters can then be seen as **emergent properties of the interaction of the three factors)** and *not* as directly predetermined by UG.

# Two third-factor principles

- **Feature economy (FE):** postulate as few formal features as possible

- **Input Generalisation (IG):**

  For a given set of features F and a given set of functional heads H, given a trigger for feature $f \in F$ of a functional head $h \in H$, the learning device generalises f to all functional heads $h_1 \ldots h_n \in H$.

# The Learning Procedure

(i) default assumption: ¬∃h [ F(h)]

(ii) if F(h) is detected, generalise F to all relevant cases

 (∃h [ F(h)]→ ∀h [ F(h)]);

(iii) if ∃h ¬[ F(h)] is detected, restrict h and go back to (i);

(iv) if no further F(h) is detected, stop.

# **Hierarchy 2: Null arguments**

Are uφ-features obligatory on all probes?

No: **Radical Pro-drop**
(least-marked option: no features)

Yes: Are uφ-features fully
specified on all probes?

Yes: **Pronominal-arguments**
(next least-marked)

No: Are uφ-features fully
specified on some probes?

No: **Non-null-subject**
(feature economy, given some features)

Are the uφ-features of some
specific (set of) head(s) {T,
v,…}fully specified?

Yes: **Italian, etc. ..**

And so on down
to microparameters…

# The third-factor principles

- lead all the relevant functional heads to prefer to "point the same way".

- NOT grammatical principles, but acquisition strategies (deriving from computational conservatism of the learning device)

- Parameters, in all their forms, are emergent properties deriving from the interaction of the three factors.

# Moving "**down the hierarchies**":

- ☐ Systems become more marked
- ☐ Parameters move from meso to micro to nano
- ☐ Parameters become intrinsically more complex, having a longer description (the conjunction of all the "dominating nodes")
- ☐ Parameters are further along a learning path
- ☐ And are diachronically more unstable

# No-choice parameters

- **Biberauer, Holmberg, Roberts & Sheehan (2010) and Biberauer, Roberts & Sheehan (2013):** UG in fact makes available certain formal underspecified options which can only be set one way

- **no-choice parameters** which are always set a given way *because of functional pressures*

- see also Baker (1996:128)

# Hierarchy 3: word structure

Given that UG allows (defective) goals to incorporate into their probes as a special case of Agree (see Roberts 2010), but doesn't mind which goals do this:

a. Do some probes trigger head-movement?
   N: **analytic** (FE & IG)

b. Do all probes trigger head-movement?
   Y: **polysynthesis** (IG)

c. Y: does some specific (subset of) {C, T …} trigger head-movement ? And so on …

# Inversion (T-to-C movement)

- ☐ **Mesoparameter:** "full" V2 (all non-Modern English Germanic)
- ☐ **Microparameter:** "residual" V2 (i.e. interrogatives, conditionals, optatives)
- ☐ **Nanoparameter:** contemporary English conditional and optative inversion

**(Biberauer & Roberts 2014)**

# Hierarchy 4: A'-movement

- ☐ Three main "edge positions": left periphery (CP), *Mittelfeld* (vP) and the nominal periphery (DP);

- ☐ Variation in which positions can be targetted/function as "escape hatches";

- ☐ These ideas can be technically expressed in terms of Edge Features and phases.
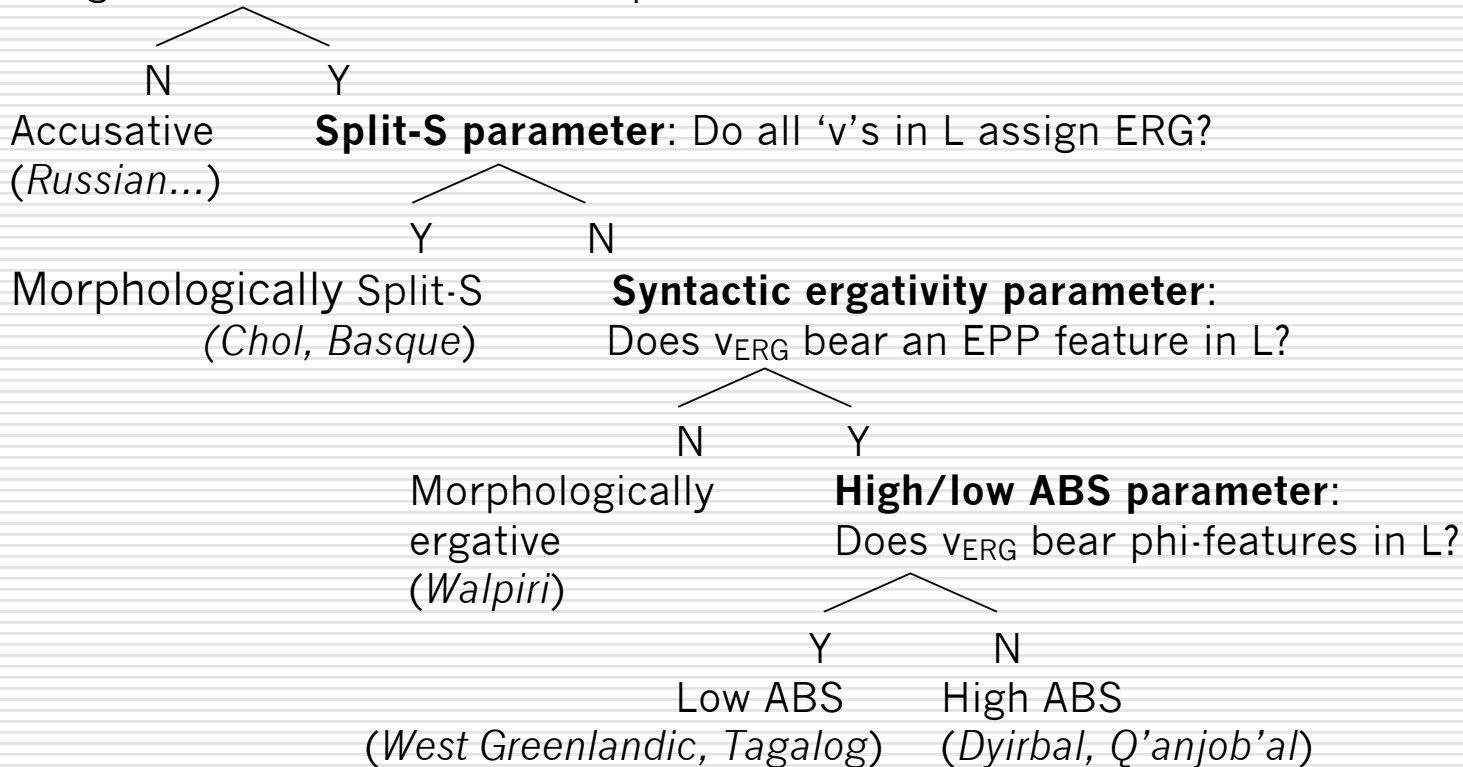
# Variation in A'-movement

☐ Maximal case: movement to all three edges, including subextraction from DP (giving rise to surface separation of adnominal modifiers from N): Latin (Ledgeway 2012), most Slavonic, Warlpiri;

☐ Minimal case: no overt wh-movement, highly restricted topicalisation, no scrambling (Mandarin?);

☐ Intermediate cases: Japanese, German, Romance, English …

# Hierarchy 5: Case/alignment
## (Sheehan 2013)

**Basic alignment parameter**: Does transitive 'v'
assign theta-related ERG to its specifier in L?

N   Y

Accusative  **Split-S parameter**: Do all 'v's in L assign ERG?
(*Russian…*)

     Y   N

Morphologically Split-S  **Syntactic ergativity parameter**:
 (*Chol, Basque*)  Does $v_{ERG}$ bear an EPP feature in L?

        N   Y

    Morphologically  **High/low ABS parameter**:
    ergative    Does $v_{ERG}$ bear phi-features in L?
    (*Walpiri*)

         Y   N

      Low ABS  High ABS
   (*West Greenlandic, Tagalog*) (*Dyirbal, Q'anjob'al*)

# Basic alignment

**Accusative (ACC)**: English (Germanic)

(1)     He is kissing her

        She is eating

**Ergative (ERG)**: Yup'ik (Eskimo-Aleut) – Bobaljik (1993: 3)

(2)     Angute-m qusngiq          ner-aa.
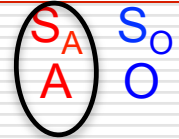        man-ERG reindeer.ABS      eat-TR.3s/3s

        'The man is eating (the) reindeer.'

(3)     Qusngiq                   ner' -uq.

        reindeer.ABS  eat-INTR.3s

        'The reindeer is eating.'

# Split-S alignment

(Western) Basque, SOV (Laka 2006):

$S_A$  $S_O$
$A$  $O$

(1)  Txalup-a  hondora-tu  da.
  boat-DET.ABS  sink-PERF  is
  'The boat sank.'

(2)  Ekaitz-a-k  txalup-a hondora-tu  du.
  storm-DET-ERG boat-DET.ABS sink-PERF has
  'The storm sank the boat.'

(3)  Oli-k  lo  egi-ten  du.
  Oli-ERG  sleep  do-IMPF  has
  'Oli sleeps.'

# Consequences of the approach

- ☐ Reconciliation of micro- and macro-approaches
- ☐ Reduction of innate endowment (no UG parameters)
- ☐ Enhancement of micro analyses with no loss of explanatory adequacy
- ☐ Reduction of variation space
- ☐ Prediction of learning paths

# The learning path

- All other things being equal, the networks predict the learning path. More interestingly, they can be thought of as forming "epigenetic landscapes" down which the learner charts its path until it comes to a natural resting point, i.e. when it reaches a "terminal node". Since the learner wants to stop as soon as it can (conservatism), higher terminal nodes represent more highly valued systems. Many clear predictions here for the relations between acquisition, typological skewing and change.

# **Possibilities opening up**

- ☐ Typological profiling of languages/ families

- ☐ Comparison of  microvariation across families

- ☐ A new take on diachronic (in-)stability

# Questions

- ☐ Form of the hierarchies: binary-branching? deterministic? Interconnected?

- ☐ How many hierarchies are there?

- ☐ Exactly which features vary, and why?

- ☐ Interaction with "spell-out", i.e. PF variation.

# Extending Emergentism

☐ Classical P&P approach (both GB and Minimalist, see Chomsky 1981, 1995): parameters and features UG-given (Factor One);

☐ Conservative emergentist view (see above): parameters are emergent;

☐ Radical emergentist view (Biberauer 2013): features are emergent.

# Formal features as emergent properties

□ UG: merely specifies [uF] vs [iF] (or [Att:val] where "val" can be blank);

□ PLD: drives postulation of set of features by learner;

□ F3: Feature Economy and Input Generalisation as above, interact with each other, UG and PLD so as to create hierarchies.

# **The basic learning path I**

- ☐ Assume no F (FE, IG, but usually not a choice);

- ☐ Generalise F triggered by PLD (IG): favours macroparameters;

- ☐ Retreat from generalisation triggered by PLD;

- ☐ Make a featural distinction and do it all again (move down the hierarchy).

# The learning path II

☐ Another way to look at it:

i. Postulate NO formal features on heads;

ii. Postulate ALL heads have a given feature;

iii. Postulate SOME heads have the feature (i.e. make a new categorial distinction).

# Complexity (Biberauer, Holmberg, Roberts & Sheehan 2014)

☐ As we move downwards along the hierarchies, parameters become more "micro" (going from macro to meso to micro to nano);

☐ The lower parameters behave in a non-uniform, differentiated fashion which is inherently more complex than the systems defined higher up.

# Measuring syntactic complexity

☐ All else being equal, there should be a roughly 50/50 chance of a given choice at each independent choice point, making lower positions in the hierarchy cumulatively less probable. The probability associated with a given output of the hierarchy is $0.5^n$ (where n=the level of embedding in that hierarchy)

# Arising issues:

- ☐ **Potentially, no system can be maximally unmarked, i.e. "at the top" of all 5 hierarchies**

- ☐ the options at the tops of the hierarchies are partially incompatible, so there can be no language which systematically reflects all and only these options.

# Properties of the least-marked possible system:

- ☐　　Harmonically head-initial;
- ☐　　Discourse pro-drop;
- ☐　　"Deep" analyticity;
- ☐　　No A'-movement (?);
- ☐　　No A-movement (??).

☐ no mechanism of focussing, topicalisation, wh-movement, scrambling, etc),

☐ rigid (Comp) > (Aux) > Neg > Adv > SVO order in the clause.

☐ We conjecture that such a system is unattested.

# A thought experiment

☐ Applying the formula discussed above, as we go down the hierarchy the probability of being assigned a given parameter value decreases as a function of "depth $p = 0.5^n$ (where n= level of embedding).

☐ the overall probability of a set of parameter values across the hierarchies can then be calculated.

# *English*

- ☐ harmonically head-initial (0.5 on Hierarchy 1),

- ☐ non-pro-drop (0.125 on Hierarchy 2),

- ☐ shows Aux but not V-movement (0.03125 on Hierarchy 3),

- ☐ has *wh*-movement but no scrambling (0.03125 on Hierarchy 4),

- ☐ and is accusative (0.5 on Hierarchy 5).

# *Mohawk (Baker 1996)*

- ☐ head-initial (by default)
- ☐ has pronominal arguments (0.25),
- ☐ polysynthetiic (0.25)
- ☐ free word order (0.25)
- ☐ split-S alignment (0.25).

- ☐ The product of these probabilities is 0.195%.

# *Mandarin Chinese*

- ☐ harmonically head-final in [+N] but not in [+V] (0.0625),

- ☐ radical pro-drop (0.5),

- ☐ highly analytic (0.5),

- ☐ has topicalisation to the left-periphery scrambling, and no *wh*-movement (0.125),

- ☐ accusative (0.5) → overall probability for this language is 0.098%

# *Japanese (and Korean?)*

- ☐ harmonically head-final (0.25),
- ☐ radical pro-drop (0.5),
- ☐ agglutinating in both verbal and nominal domains (0.5),
- ☐ *wh*-in-situ + scrambling (0.125)
- ☐ accusative (0.5).
- ☐ This gives an overall complexity index (probability) of 0.391%

# *Basque*

- ☐ is harmonically head-final (0.25),
- ☐ has pronominal arguments (0.25),
- ☐ is agglutinating (0.5),
- ☐ has *wh*-movement+scrambling (0.125)
- ☐ and split-S alignment (0.25),
- ☐ complexity index: 0.098%, identical to that of Mandarin Chinese.

# Relative complexities in terms of probabilities

- Japanese/Korean: 0.391%
- Mohawk: 0.195%
- Mandarin: 0.098%
- Basque: 0.098%
- English: 0.003%


- The lower the number, the lower the probability ($\rightarrow$ greater complexity)

# **Three comments**

- ☐ English is the most complex (NB the effect of the auxiliary system)
- ☐ No extreme outliers
- ☐ Japanese/Korean least complex

# Macro-Altaic?

- harmonically head-final word order, radical pro-drop, agglutinating morphology, *wh*-in-situ with scrambling and accusative alignment are properties of a large number of languages spoken across the northern part of Asia (Macro-Altaic: Miller 1971; Menges 1975).

# **Conclusion to Part One**

- ☐ A new approach to P&P
- ☐ Radically minimalist (very little in UG, nothing about parameters)
- ☐ Implications for typology, acquisition and diachrony
- ☐ Interactions with FOFC (which constrains the word-order hierarchy)
- ☐ A new approach to complexity.