Syntactic universals in the lab: New methods and approaches

Jennifer Culbertson, University of Edinburgh

Advanced Core Training in Linguistics (ACTL), Summer 2015.
University College London.

Lecture 1: Theoretical & empirical issues, early ALL studies

Opening question: What are "universals" in syntax?

The draw of typological universals

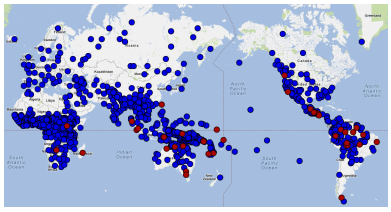


Figure 1: Greenberg's Universal 1: S before O in declaratives.

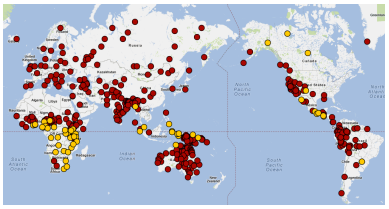


Figure 2: Suffixing preference.

Constrained diversity? There is considerable diversity across languages, yet the existence of typological generalizations—often called universals—suggests that variation is constrained...Very tantalizing, seems like something we should explain...

Constrained acquisition? We know that inductive biases are necessary for language acquisition. Arguments from the poverty-of-the-stimulus have traditionally been used to suggest these need to be extensive and specific to language. Typological universals might be a reflection of this.

Major theoretical/empirical issues arising from typological universals

1. Types of universals

- Hard vs. soft
  - Some universals are (claimed to be) absolute or exceptionless (1,3)
  - Others are (claimed to be) tendencies (2,4).

1. All languages have consonants/oral vowels/pronouns/recursion/wh-movement.
2. Languages tend to have nasal consonants. (555/567 = .98)[1]
3. If 1st/2nd person distinct reflexive pronouns, then 3rd person. [unidirectional]
4. If SOV then postpositions. [bidirectional] (472/486 = .97)[2]

- Formal vs. substantive[3]
  - Formal = properties of the grammatical architecture, types of rules, interaction.
  - Substantive = related to categories or specific rules the grammar uses

Which types? Syntacticians typically interested in **formal, absolute** universals, but many **substantive** universals appear to be **soft**.

- Do we want to explain hard *and* soft universals?[4,5]
- If so, how can we deal with exceptions while still explaining what look like interesting constraints on linguistic structure.

1. Status quo: **Panini's chakra** plus some notion of markedness.[6,7]
2. Soft cognition-general (substantive) biases shaping learning and use.
3. Probabilistic constraints in the theory itself.
4. (2) and (3).[9]

[1] I. Maddieson, *The World Atlas of Language Structures Online* (2013)

[2] M. Dryer, *The World Atlas of Language Structures Online* (2013)

[3] N. Chomsky, *Aspects of the Theory of Syntax* (MIT Press, Cambridge, MA, 1965)

"...Greenberg's universals are really surfacing properties of language that typically can be explained in functionalist terms and allow for...exceptions....they are basically tendencies...but that is not the kind of thing that generative grammarians have focused on..."

C. Boeckx pp. 195–199 (2009)

[4] F. Newmeyer, *Possible and Probable Languages* (OUP, 2005)

[5] R. Bod, J. Hay, S. Jannedy, eds., *Probabilistic Linguistics* (MIT Press, 2003)

[6] A. Prince, P. Smolensky, *Science* **275**, 1604 (1997)

[7] G. Cinque, *Linguistic Inquiry* **36**, 315 (2005)

[8] B. Hayes, C. Wilson, *Linguistic Inquiry* **39**, 379 (2008)

[9] J. Culbertson, P. Smolensky, C. Wilson, *Topics in Cognitive Science* **5**, 392 (2013)

2. Interpreting typological data

- What shapes typology?[10,11]
  - accidental factors (e.g. guns, germs and steel)
  - genetic relationships among languages
  - geography/contact
  - cognition-general and -external constraints or biases
  - domain-specific linguistic constraints or biases
- How do we interpret the data then?[12]
  1. **Sampling.** Which languages to use? Convenience sampling (no restrictions)? Stratified probability sampling (control influence of known factors, e.g. genetic and geographic distance)?[13] Are there enough languages to make inferences about possible/impossible languages?[14]
  2. **Establishing types.** Categorical vs. continuous definition of language types? E.g. in case-marking systems, word order classification.
  3. **Interpreting variation.** What to do with exceptions? Error (probability) samples should balance out, so by "correcting" non-accidentally chosen errors in uncommon types, we reduce the validity of the sample. However, non-occurrence of particular types is not necessarily meaningful anyway (i.e. building into the theory amounts to over-fitting the data).

WALS

Some of these are of more interest than others to cognitive scientists...NB: recent high-profile articles have argued against cognitive/linguistic universals.

[10] N. Evans, S. C. Levinson, *Behavioral and Brain Sciences* **32**, 429 (2009)

[11] M. Dunn, S. Greenhill, S. Levinson, R. Gray, *Nature* **473**, 79 (2011)

[12] M. A. Cysouw, *Quantitative Linguistics: An International Handbook* (Mouton de Gruyter, Berlin, 2005), pp. 554–578

[13] J. Rijkhoff, D. Bakker, K. Hengeveld, P. Kahrel, *Studies in language* **17**, 169 (1993)

[14] S. T. Piantadosi, E. Gibson, *Cognitive Science* **38**, 736 (2014)

$A \rightarrow B$?

	A	¬A
B	14	72
¬B	0	14

	A	¬A
B	26	48
¬B	0	26

3. Domain of universals in cognition

- Generative view
  - Principles & Parameters: macro-parameters (head-directionality, null-subjects) constrain acquisition/typology
  - Minimalism: movement away from explaining "surface-y tendencies", instead have few, very general principles (e.g., Merge, Agree, efficient computation), with variation in other components (morphology), or extra-linguistic
- Functionalist view: processing-related biases shape usage and thus typology
- Psychologist view: general pattern learning mechanisms along with constraints on memory/encoding

4. Role of adults vs. children in shaping typology

- Errors or reanalysis during learning lead to cross-generational changes (fast or slow)[15,16]
- Inductive biases accrue over generations[17]
- Innovation and accommodation by adult speakers[18,19]

These issues...motivate the search for converging empirical evidence and models of (potentially) soft biases, explorations of their domain of application, and comparisons across developments.

[15] D. Lightfoot, *Lingua* **100**, 171 (1997)

[16] C. D. Yang, *Language variation and change* **12**, 231 (2000)

[17] S. Kirby, H. Cornish, K. Smith, *Proceedings of the National Academy of Sciences* **105**, 10681 (2008)

[18] W. Labov, *Principles of language change, Vol. 1: internal factors* (Blackwell, New York, 1994)

[19] P. Auer, F. Hinskens, *Dialect Change*, P. Auer, F. Hinskens, P. Kerswill, eds. (Cambridge University Press, New York, 2008), pp. 335–357

Progress, using Artificial Language Learning

WHY ALL? Converging empirical evidence comes from multiple sources

1. From state transitions (aka, **language change**)[20]
   - Same issues as synchronic typological data (e.g., need lots!)
2. From **natural language acquisition**
   - Subtle biases not necessarily evident in production data
   - Many confounding factors
     - Each learner's experience (i.e. input) is different
     - Frequencies of features/structures cannot be controlled
     - Impossible to compare attested vs. unattested structures
3. From **artificial language learning**
   - Can get lots of data
   - Established methods of statistical testing
   - Strict control over input (removal of confounding factors)
   - Can ask questions not easily testable using natural language learning data:
     - Why do languages have certain features but not others?
     - What (if any) prior knowledge or biases do learners bring to the task? Do these align with typology?
     - What information can/will learners use?
     - How abstract/generalizable is the knowledge speakers have about language?
     - How domain-specific are the mechanisms of or constraints on language learning?
     - How species-specific are the mechanisms of or constraints on language learning?

[20] M. Dunn, S. Greenhill, S. Levinson, R. Gray, *Nature* **473**, 79 (2011)

WHAT ARE THE POTENTIAL LIMITATIONS OF ALL (FOR SYNTAX?

- Need semantic reference
  - Contrast with ALL studies of phonetics/phonological → added complexity
- Need an artificial lexicon
  - Even though, that's not really what we care about
  - Actually, not clear we do...[21,22]
- Need child learners?
  - Depends in part on who we think influences language change most
  - We'll spend a lecture talking about this[23,24]
- Imagination...
  - Syntactic patterns can be abstract, involve complex structures
  - How do we get people to learn those?
  - How can we be sure they are actually learning what we think they are?

[21] K. Smith, E. Wonnacott, *Cognition* **116**, 44 (2010)

[22] J. Culbertson, D. Adger, *Proceedings of the National Academy of Sciences* **111**, 5842 (2014)

[23] C. Hudson Kam, E. Newport, *Cognitive Psychology* **59**, 30 (2009)

[24] J. Culbertson, E. L. Newport, *Cognition* **139**, 71 (2015)

Classic artificial grammar/language learning studies

NB: starting from very general properties of language, moving to more specific.

Reber: AG rule learning[25]

- Research question: how do people become (implicitly) sensitive to statistical variation in a set of stimuli?
- Procedure:
  - 2 Artificial grammar learning experiments, introduced as "memory-task"
  - Participants had to learn 28 strings ("sentences"), seen in sets of four
  - After each set of four sentences, recall test (with uninformative feedback)
  - In Exp. 2 forced-choice testing phase with new exemplars
- Manipulation:
  - Grammar condition
  - Random condition
- Results/Conclusion:
  - Both conditions improved over 7 blocks, but grammar condition more so.
  - Participants in Exp. 2 generalized to novel strings.
  - People can do perceptual learning of grammar-like rules.

[25] A. S. Reber, *Journal of Verbal Learning and Verbal Behavior* **6**, 855 (1967)

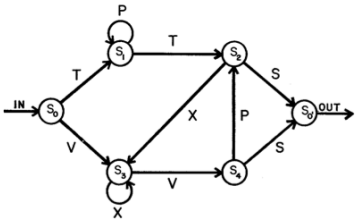


Figure 3: Reber finite-state grammar

THE FREQUENCY WITH WHICH GRAMMATICAL (G) AND NONGRAMMATICAL (NG) RESPONSES WERE MADE TO GRAMMATICAL AND NONGRAMMATICAL ITEMS

	Response	
Item	G	NG
G	345	95
NG	91	349
Total	436	444

Figure 4: Reber forced-choice results.

Saffran et al: word segmentation from statistics[26,27]

- Research question: are young children sensitive to differences in transitional probabilities among syllables?
- Procedure:
  - Artificial grammar passive listening task
  - 8-month-olds heard continuous stream of syllables containing 4 "words" (*bidaku, tupiro, golabu, padoti*)
  - Listening times to words, non-words, part-words measured
- Manipulation:
  - High transitional probability (e.g., $p(da|bi) = 1$)
  - Low transitional probability (e.g., $p(tu|ku) = 0.33$)
- Results/Conclusion:
  - Words distinguished from non-words and part-words (novelty effect).
  - Further study de-confounded TP and frequency, replicated result.
  - This mechanism is available, may be used during learning.

[26] J. R. Saffran, E. L. Newport, R. N. Aslin, *Journal of Memory and Language* **35**, 606 (1996)

[27] R. N. Aslin, J. R. Saffran, E. L. Newport, *Psychological Science* **9**, 321 (1998)

Transitional probability (TP) of XY
$Y|X = \frac{\text{frequency of } X}{\text{frequency of } Y}$

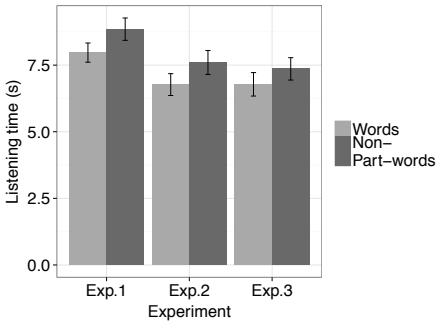


Figure 5: Results from Saffran (1996) and (1998).

Morgan et al.: cues to phrase structure[28]

- Research question: do commonly occurring cues to phrase structure (prosody, function words, morphological dependencies) facilitate learning?
- Procedure:
  - Artificial language learning task, semantically meaningful language
  - 4 blocks of training and testing, listened to and saw sentences in the language
  - Tested on vocabulary learning, rules of linear order, and "constituents" (sentence fragments)

[28] J. L. Morgan, R. P. Meier, E. L. Newport, *Cognitive Psychology* **19**, 498 (1987)

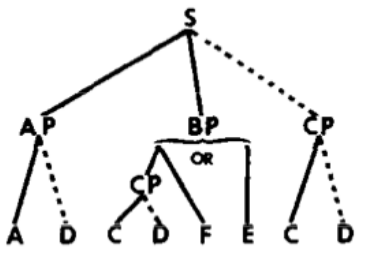
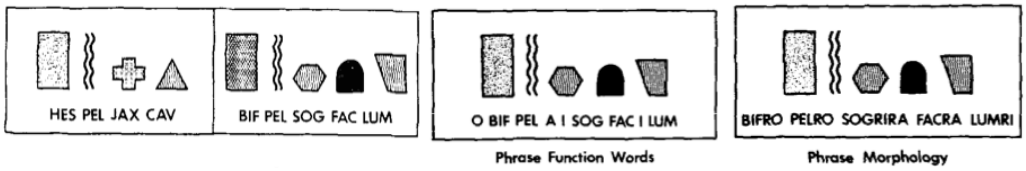


Figure 6: Morgan et al. phrase-structure grammar.



[A D] [E] [D] [A D] [C F] [C] [o AD] [a [i CD] F] [i CD] $[A_i\ D_i][[C_{jk}\ D_{jk}]\ F_k][C_j\ D_j]$

- Manipulation:
  - Grouping cue aligning with phrase structure
  - Grouping cue not aligning with phrase structure
  - No grouping cue
- Results:
  - In all cases, better learning when cues align with grammar.
  - Biggest improvement for "conditional" aspects of the language, i.e., dependencies among word classes (as opposed to "unconditional" aspects like every phrase must begin with a single "A").
  - These properties of language may be present *because* they aid learning.

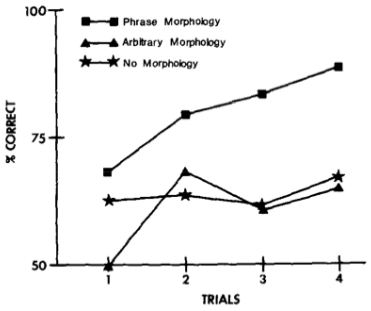


Figure 7: Morgan et al. Exp.1 results.

Smith et al.: Possible vs. impossible rules[29]

- Research question: to what extent can a polyglot savant learn "impossible" rules.
- Procedure:
  - Explicit teaching with translation and sentence correction tasks
- Manipulation:
  - Berber
  - Artificial language Epun with impossible rules
- Results/Conclusion:
  - Berber:
    - Very good learning of inflectional morphology
    - Assumed prepositions (correct), but SVO (incorrect) despite VSO data
    - Accepted null-subjects (correct) without any data
    - Translated *that*-trace violations as extracted objects
  - Epun:
    - Could not learn negation, past tense, or emphatic
    - No problem with conjunct agreement
    - Control subjects learned negation and past tense (not the others)
  - No evidence for parameter-resetting view of L2-acquisition (just transfer)
  - They argue that "current linguistic theory makes plausible predictions about what constitutes a possible human language."
  - Conceptual replication[30] using German speakers learning actual vs. "impossible" rules in Italian/Japanese learn all rule types, but Broca's area activation correlated with accuracy only for actual rules.

[29] N. V. Smith, I.-M. Tsimpli, J. Ouhalla, *Lingua* **91**, 279 (1993)

The polyglot savant is **Christopher** General impairment (he was institutionalized), but spared language abilities, in fact extremely talented at learning/translating languages. (Used as evidence of modularity of language by these authors in earlier work).

(1) S-V inversion for negation
(2) Object-fronting for past tense
(3) Emphatic attached to 3rd word
(4) strange conjunct agreement (1st+3rd=2nd)

"Who did you say that knows Munat?" → "...that Munat knows?"

[30] M. Musso, *et al.*, *Nature neuroscience* **6**, 774 (2003)