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 2: Core results illustrating methodology

Overview of methods in ALL for syntax

Methods to cover Several quite distinct methods have been used to investigate typological universals in syntax using artificial language learning.

1. Basic ease of learning
2. Iterated learning paradigm[2]
2. Mixture-shift paradigm[1]
4. Poverty-of-the-stimulus paradigm[3]

[1] C. Hudson Kam, E. Newport, *Language Learning and Development* **1**, 151 (2005)

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

[3] C. Wilson, *Cognitive Science* **30**, 945 (2006)

Regularization

Conditioned vs. unconditioned variation. Examples of conditioned variation? How about unconditioned? How do people treat unconditioned variation?

- Unconditioned (random) variation is rare
- Present in second language learner speech, pidgins/early creoles, but...
  - Regularization in process of creolization[4]
  - Regularization of late-learner input by children[5]
  - Regularization (aka maximization) in non-linguistic probability learning tasks[6]

Regularization in ALL (by adults)[7]

- Research question: do adults regularize inconsistent (random) variation?
- Design of the language
  - Lexicon: 12 verbs, 36 nouns (2 classes), 2 determiners, negative marker
  - Grammar: (Neg)VSO word order
  - Stimuli: female speaker describing movies with puppets
- Procedure: 9 sessions, 30 minutes each (9-12 days)
  - Training: 8 sessions (subset of 230 sentences, each hear 4x)
  - Testing: 1 session
    - Vocabulary: provide name for object (x12)
    - Sentence completion: see scene, hear verb, provide complete sentence (x24)
    - Determiner judgment: hear sentences, provide judgment (x48)
    - General grammar: listen to pair of sentences, choose grammatical one (x16)
- Manipulation
  - Amount of complexity (=random variation in determiner choice)
- Participants: 50 native English-speaking undergraduates
- Results
  - Probability matching in presence/absence case, but...
  - *"[R]egularization behavior can be induced in adult language learners when they are given input that contains what we have called 'scattered inconsistency'."* → Increasing regularization as complexity increased

[4] G. Sankoff, *The Genesis of Language*, K. C. Hill, ed. (Karoma Publishers, Ann Arbor, MI, 1979), pp. 23–47

[5] J. L. Singleton, E. L. Newport, *Cognitive Psychology* **49**, 370 (2004)

[6] W. K. Estes, *American Psychologist* **19**, 16 (1964)

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

Figure 1: Screen shot of puppets.

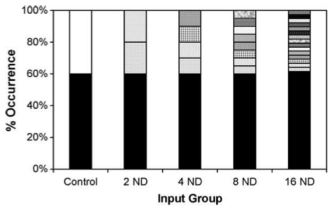


Figure 2: Determiner manipulation.

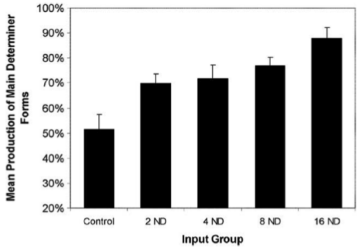


Figure 3: Average main Det production.

ITERATED-LEARNING. "Iterated learning describes the process whereby an individual learns their behaviour by exposure to another individual's behaviour, who themselves learnt it in the same way. It can be seen as a key mechanism of cultural evolution."[8]

[8] S. Kirby, T. Griffiths, K. Smith, *Current opinion in neurobiology* **28**, 108 (2014)

ITERATED REGULARIZATION[9]

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

- Research question: Are phenomena like regularization in part the result of weak biases amplified over generations of learners?
- Design of the language
  - Lexicon: 4 English nouns (cow, pig, giraffe, rabbit), 1 verb ("move"), 2 nonce plural morphemes (*fip*, *tay*)
  - Grammar: V S
  - Visual stimuli: still images with arrow indicating motion
- Procedure
  - Transmission chain: output of generation 1 used as input to generation 2, etc.
  - Training
    - Noun familiarization to ensure correct labels
    - Sentence learning (x96)
  - Testing: sentence completion (typing; x32)
- Manipulation
  - Initial chain input has two markers, used 75% and 25% respectively
  - Majority marker is counterbalanced across chains
- Participants: 65 native English-speaking undergraduates
- Results
  - *"[P]articipants copy the proportion of marking that they see, and that proportion of marking is (on average) preserved across all five participants in a chain."* → On average and in initial generation, (statistically) matching input
  - *"[T]he elimination of unpredictable variation is cumulative, rather than purely a consequence of the behaviour of the first learner in each chain."* → Over generations, chains converge on (different) regular systems

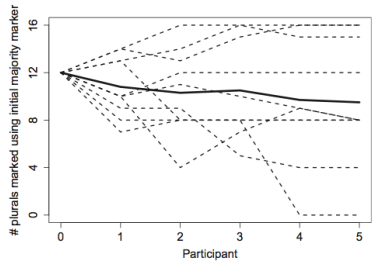


Figure 4: Number of plurals marked with majority marker on average and within individual chains.

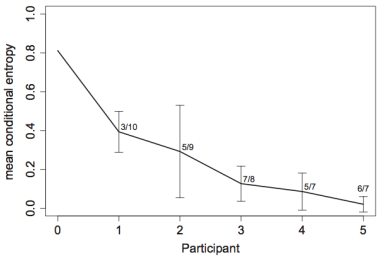


Figure 5: Reduction in entropy over generations (and number of languages per chain with variation that is significantly non-random).

TAKE-HOME MESSAGE: REGULARIZATION.

- Illustrates two methods
  - Mixture-shift (regularization) paradigm. Pros? Cons?
  - Iterated-learning paradigm. Pros? Cons?
- *Results:* gradual regularization by learners reduces unconditioned variation over generations.
- *Remaining issues:* domain-specific (part of the grammar) or general (operating for any kind of probability-learning)? adults or children?

Harmony

HARMONY. Typically called 'consistent-headedness': phrases tend to be either consistently head-initial or head-final (not both). This has a long history in syntax.

- Greenberg's Universals 2–5, a.o.[10]
- Theoretical syntacticians have codified this in various ways[11,12,13]
- Attempts to validate this tendency statistically reveal mixed results[14,15]

HARMONY IN ALL[16]

- Research question: does ease of learning parallel typological preference for consistent head ordering?
- Design of the language
  - Lexical items:
    - Single consonant instantiating each category
    - (X = Gen; Z = P; Q = N_{pl}; V = N_{sg}; S = V_{sg}; M = V_{pl})
    - e.g. N_{sg} P N_{sg} Gen N_{pl} V_{pl} = V Z V X Q M
- Procedure
  - Training: see and retype string (30 total)
  - Testing: y/n classification of strings (ungrammatical strings had single interior letter changed)
- Experimental manipulation (6x2 design)
  - Consistent vs. non-consistent head order
- Participants: 40 native English-speaking undergraduates
- Results
  - Better classification of ungrammatical items in consistent condition
  - No difference for grammatical items
  - *"[B]asic word order universals (head-ordering) can be explained in terms of non-linguistic constraints on sequential learning and processing, rather than as a product of innate linguistic knowledge."* → Hmm...really?

[10] J. Greenberg, *Universals of Language* (1963), pp. 73–113
U3: VSO → preposition
U4: SOV → postpositions
U5: SOV, N-Gen → N-Adj

[11] M. Baker, *The atoms of language: The mind's hidden rules of grammar* (2001)

[12] N. Chomsky, *Theory of markedness in core grammar* (1981), pp. 123–146

[13] L. Travis, Parameters and effects of word order variation, Ph.D. dissertation, MIT (1984)

[14] M. Dryer, *Language* **68**, 81 (1992)

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

[16] M. H. Christiansen, *The Evolution of Language: 3rd International Conference* (2000), pp. 45–48

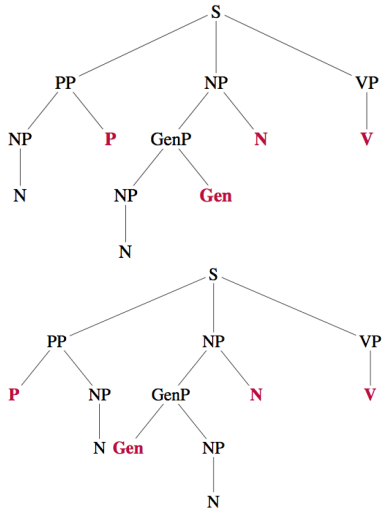


Figure 6: Consistent and mixed grammars.

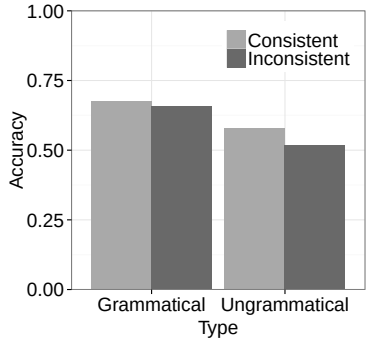
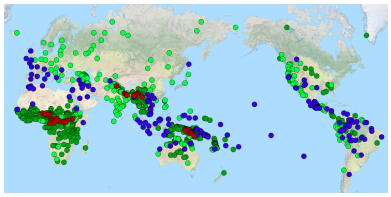


Figure 7: Classification accuracy.

Universal 18

UNIVERSAL 18. Illustrates **harmony** and apparent **asymmetry between two non-harmonic patterns**. Also illustrates recurring problem of statistic tendencies...How to test experimentally...?



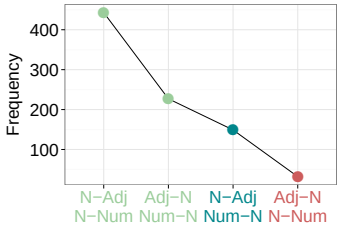


Figure 8: U18: Adj-N → Num-N.

U18 IN ALL (MIXTURE-SHIFT PARADIGM)[17]

- Research question: Do differences in regularization reveal evidence of preferences/biases parallel to Universal 18 (pro-harmony, anti-(Adj-N, N-Num)?
- Design of the language
  - Lexicon: 10 nouns, 6 adjectives, 6 numerals
  - Auditory stimuli: artificially generated
  - Visual stimuli: still images of novel objects
- Procedure: Single session (45 minutes)
  - Training:
    - Learn nouns (x50, 75% correct required)
    - See/listen to {N, Mod} phrases (x80)
    - Picture matching {N, Mod} phrases (x80)
  - Testing:
    - Produce description of picture (x80)
- Manipulation
  - Variation:
    - Dominant order (70% of utterances)
    - Unconditioned variation in order (30% of utterances)
  - Pattern:
    - Harmonic: (Adj-N, Num-N) or (N-Adj, N-Num)
    - Non-harmonic: (N-Adj, Num-N) or (Adj-N, N-Num)
    - Random condition (50% each order)
- Participants: 65 native English-speaking undergraduates
- Results
  - Most regularization of harmonic patterns
  - Least regularization of (Adj-N, N-Num)
  - Shifting of individuals toward preferred patterns
  - *"If biases in the cognitive system influence language acquisition, pushing learners in certain directions rather than others, then, all else equal, over generations of learners, languages which (better) satisfy those biases are expected to outnumber those which do not."*

[17] J. Culbertson, P. Smolensky, G. Legendre, *Cognition* **122**, 306 (2012)

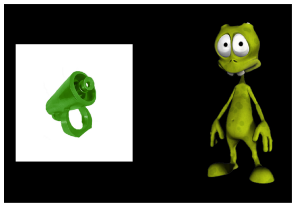


Figure 9: Example trial.

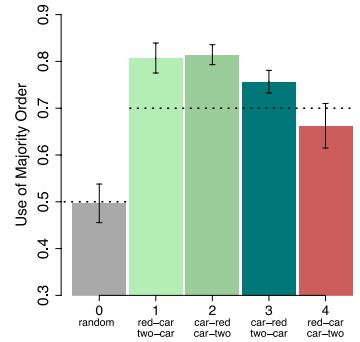


Figure 10: Use of dominant order.

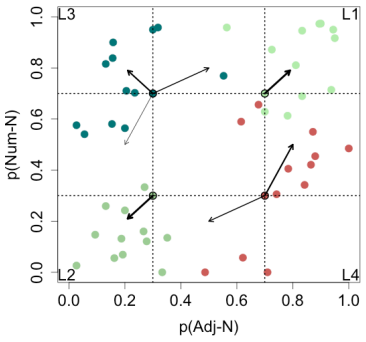


Figure 11: Individual outcomes.

TAKE-HOME MESSAGE: HARMONY AND U18

- Illustrates two methods
  - Basic ease of learning. Pros? Cons?
  - Mixture-shift (regularization) paradigm. Pros? Cons?
- *Results:* harmonic patterns are easier to learn, more likely to be regularized. Dispreferred patterns are shifted toward harmonic (linkage with change).
- *Remaining issues:* domain of bias? child vs. adults?

Universal 20

UNIVERSAL 20. Original formulation by Greenberg, but continued progress in typology resulted in several reformulations. Again illustrates clear issue of statistical tendencies *and* how to treat unattested patterns. Will return to theoretical implications and accounts in Lecture 5.

- Greenberg (1963): If prenominal Dem-Num-Adj.
  If postnominal Dem-Num-Adj or Adj-Num-Dem.
- Focus here on asymmetry among two post-nominal harmonic patterns
- *Core idea:* semantic scope (composition) → surface order

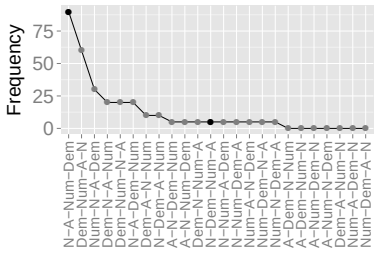


Figure 12: Pattern frequency; N-A-Num-D and N-D-Num-A highlighted.

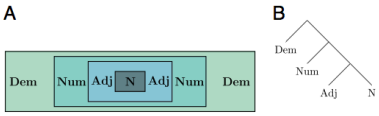


Figure 13: Illustrations of semantic scope/composition.

UNIVERSAL 20 IN ALL (POVERTY-OF-THE-STIMULUS PARADIGM)[18]

- Research question: Do learner follow surface statistics of English or more abstract knowledge about semantics-syntax mapping when inferring a new word under?
- Design of the language
  - Lexicon: 30 nouns, 10 adjectives, 10 numerals, 4 demonstratives *all English*
  - Visual/auditory stimuli: orthographic and auditory (artificially generated) presentation
- Procedure
  - 10-15 minutes total
  - Participants trained on ambiguous subset of phrases, noun + post-nominal modifier, relative order of modifiers is held out.
  - Training: see English phrase, hear translation, click on matching phrase
  - Testing: see English phrase, choose translation
- Manipulation
  - Experiment 1:

Combination	Training	Testing
{Adj, Dem}	N-Adj, N-Dem	{N, Adj, Dem}
{Num, Dem}	N-Num, N-Dem	{N, Num, Dem}
{Adj, Num}	N-Adj, N-Num	{N, Adj, Num}

  - Experiment 2:

Combination	Training	Testing
{Adj, Num, Dem}	N-Adj, N-Num, N-Dem	all combos of two mods
{Adj, Num, Dem}	N-Adj, N-Num, N-Dem	{N, Adj, Num, Dem}

- Participants: 160 Amazon Mechanical Turk workers, 32 per condition ($0.50-0.75)
- Results
  - Across both experiments, participants chose scope-isomorphic order
  - In Exp 1, strength of preference moderated by combination (scopal distance)
  - *"Learners consistently preferred the [isomorphic] order, suggesting that structural knowledge trumps distributional knowledge of English when learners make inferences about a new language system"*

TAKE-HOME MESSAGE: U20

- Illustrates poverty of the stimulus paradigm
- Conducted entirely over the web, with English lexical items
- *Results:* learners infer scope-isomorphic pattern, following typology
- *Remaining issues:* domain of bias? child vs. adults?

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

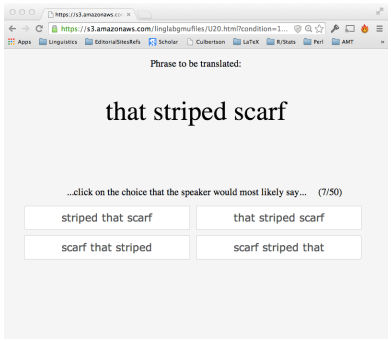


Figure 14: Example testing trial.

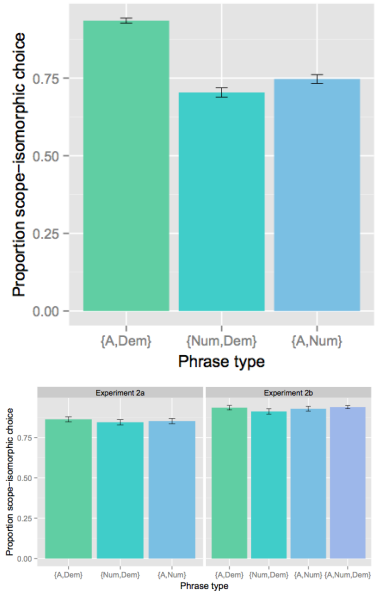


Figure 15: Average proportion scope-isomorphic choices.

Basic word order

BASIC WORD ORDER. This is Greenberg's Universal 1, plus some additional proposed tendencies relevant to the contiguity of V and O, and the general preference for SOV.[19,20]

BASIC WORD ORDER IN ALL[21]

- Research question: does ease of learning parallel relative typological frequencies of basic word orders?
- Design of the language
  - Lexical items:
    - 8 nouns (3 male, 3 female, 2 inanimate), 6 actions (2 intransitive, 4 transitive), 2 determiners
    - Words randomly chosen from pool of monosyllabic (Det), and trisyllabic (content words) per participant
  - Auditory stimuli: artificially generated individual words (concatenated for sentences)
  - Visual stimuli: videos (12 intransitive, 168 transitive)
- Procedure *"...learn an alien language..."*
  - Training: noun/action learning;
    - Heard/saw scenes (12 trials x 3 blocks)
    - Forced choice test, two minimally differing scenes, pick correct one given sentence (6 trials x 3 blocks)
  - Testing: production; see scene, construct sentence by clicking on vocal items needed to describe it.
- Experimental manipulation (6x2 design)
  - Between-subjects manipulation of word order
  - Basic: SOV, SVO, VSO, VOS, OVS, OSV
  - DP: N-Det, Det-N
- Participants: 285 Amazon Mechanical Turk workers, paid $0.75.
- Results
  - SVO > SOV > VSO > OVS > OSV > VOS
  - Clear influence of English, but SOV is next best
  - Good to have S before O? (SVO, SOV, VSO > OVS, OSV, VOS)
  - Good to be verb medial? (OVS best of the worst)

[19] M. Baker, *The atoms of language: The mind's hidden rules of grammar* (Basic Books, New York, NY, 2001)

[20] S. Goldin-Meadow, W. C. So, A. Özyürek, C. Mylander, *Proceedings of the National Academy of Sciences* **105**, 9163 (2008)

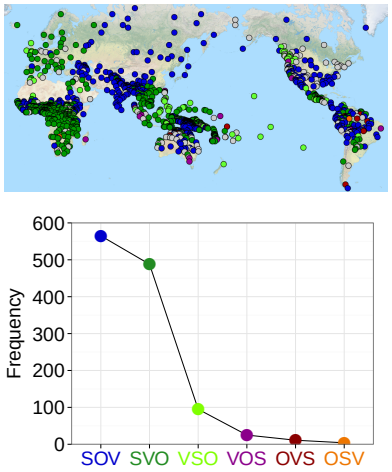


Figure 16: Asymmetric frequency in basic word order.

[21] H. Tily, M. Frank, T. Jaeger, *Proceedings of CogSci 33* (2011), pp. 1364–1369

Figure 17: Screen shot of training trial.

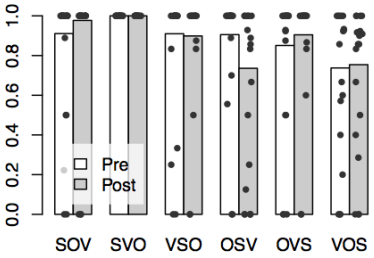


Figure 18: Production order accuracy.