The ULTRA model and Universal 20

David P. Medeiros & Ryan Walter Smith

University of Arizona

INTRODUCTION

- We present a parser-as-grammar model, the Universal Linear Transduction Reactive Automaton (ULTRA).
- ULTRA sorts word orders into a universal bottom-up order using Knuth's (1968) stack-sorting algorithm.
- Any input containing the subsequence 213 (i.e. a word order expressing a hierarchy as mid…high…low) is incorrectly sorted (and rejected).
- The model correctly predicts and interprets attested word orders, and rules out unattested word orders in natural languages in several domains, including Universal 20 (Greenberg 1963), verb clusters (Wurmbrand 2003), and basic clausal word order.
- Leads to a characterization of these universals as *213.

UNIVERSAL 20

- Previous work has uncovered several word order universals involving 213-avoidance.
- Greenberg's (1963) Universal 20.
  - "When any or all of the items (demonstrative, numeral, and descriptive adjective) precede the noun, they are always found in that order. If they follow, the order is either the same or its exact opposite." (Greenberg 1963: 87)
  - Subsequent work significantly refined the descriptive details.
- Of 24 logically possible orders of demonstrative, number, adjective, and noun, only 14 are attested (Cinque 2005; 2014).

Dem Num Adj N	Dem Num N Adj	Dem N Num Adj	N Dem Num Adj
103 = 18.6%	52 = 9.40%	22 = 3.98%	18 = 3.25%
319 = 20.8%	121 = 7.88%	38 = 2.48%	50 = 3.26%
*Num Dem Adj N	*Num Dem N Adj	*Num N Dem Adj	*N Num Dem Adj
0	0	0	0
*Adj Dem Num N	*Adj Dem N Num	Adj N Dem Num	N Adj Dem Num
0	0	5 = .90%	29 = 5.24%
		15 = .98%	78 = 5.08%
*Num Adj Dem N	Dem Adj Num N	Dem N Adj Num	N Dem Adj Num
0	24 = 4.34%	69 = 12.5%	16 = 2.89%
	36 = 2.35%	132 = 8.60%	23 = 1.50%
*Adj Num Dem N	Num Adj Dem N	Num N Adj Dem	N Num Adj Dem
0	16 = 2.89%	47 = 8.50%	25 = 4.52%
	40 = 2.61%	188 = 12.2%	41 = 2.67%
*Adj Num N Dem	*Adj N Num Dem	Adj N Num Dem	N Adj Num Dem
		11 = 1.99%	116 = 21.0%
		24 = 1.56%	430 = 28.0%

Cinque's (2014) sample. 553 genera, 1535 languages *Yellow and white cells are attested orders. Grey cells are unattested orders. The yellow cells in the top-left and bottom-right corners are the most frequent orders cross-linguistically.*

THE ULTRA MODEL

- The core of the ULTRA model is Knuth's (1968) stack-sorting algorithm, defined below.

STACK-SORTING ALGORITHM	DEFINITIONS
While input is non-empty,	I: next item in input.
If I > S, Push.	S: item on top of the stack.
Else Pop.	I > S if I is deeper than S in the base (e.g. N > Adj)
While Stack is non-empty,	Push: move I from the input onto the stack.
Pop.	Pop: move S from the stack to the output.

- Also fitted with a set of cartographies motivated in previous research, and maps each to a number.
  - Dem < Num < Adj < N (Cinque 2005)
    1 < 2 < 3 < 4
- All 213-avoiding orders are sorted into a uniform 321-like order, reflecting their order of composition.
- Applied to the DP domain, we get Cinque's typology.

1234 → 4321	1243 → 4321	1423 → 4321	4123 → 4321
2134 → *2431	2143 → *2431	2413 → *4231	4213 → *4231
*3124 → *3421	3142 → *3421	3412 → 4321	4312 → 4321
*1324 → *3421	1342 → 4321	1432 → 4321	4132 → 4321
*2314 → *3241	2341 → 4321	2431 → 4321	4231 → 4321
*3214 → *3241	3241 → *3421	3421 → 4321	4321 → 4321

Result of stack-sorting logically possible DP orders.

*213 BEYOND UNIVERSAL 20

- The ULTRA model generalizes beyond Universal 20.
- For instance, Germanic verb clusters exhibit a similar 213-avoiding profile.
  - "First, of the six possible combinations involving three verbal elements, five orders are robustly attested. One order, the 2-1-3 order, however, is highly infrequent, perhaps impossible, as an order in verb clusters […]." (Wurmbrand 2006: 14)

Language	3-2-1	3-1-2	1-2-3	1-3-2	2-3-1	2-1-3
Afrikaans	*	*	*	*	OK	*
Dutch Standard	*	*	OK	*	*	*
Dutch Dialects	OK	OK	OK	OK	*	*
Frisian	OK	*	*	*	*	*
German Standard	*	*	OK	*	OK	*
German, Austrian dialects	OK	OK	*	OK	*	*
Swiss dialects	*	OK	OK	OK	*	*
West Flemish	*	*	OK	*	OK	*

- Availability of IPP constructions in West Germanic language. Note that the 213 order is uniformly unattested.

- Similarly, in the domain of basic clause order, for languages with one dominant information-neutral word order, OSV is extremely rare, perhaps spurious.
- If SOV is taken as 123 order, we have the following (data from WALS, out of 1377 languages).
  - SOV: 565   123
  - SVO: 488   132
  - VSO: 95   312
  - VOS: 25   321
  - OVS: 11   231
  - ?*OSV: 4   *213
- This 213-avoidance in many domains falls out from the architecture of ULTRA.

PREVIOUS ACCOUNTS

- We contrast the ULTRA account of 213-avoidance with three others, focusing on Universal 20.
- Cinque's (2005) account of noun phrase order stipulates a ban on remnant movement, in conjunction with LCA-style linearization (Kayne 1994) of a Merge structure.
- Steddy and Samek-Lodovici (2011) explain the same facts in terms of harmonic bounding of 213-containing candidates under Left-Alignment constraints.
- Both accounts share three unappealing features.
  1. They invoke linearization principles distinct from the most basic mechanisms of grammar.
  2. The relevant patterns require language-particular learning (of the proper derivation-driving features, or constraint ranking).
  3. Some orders have ambiguous analyses, consistent with multiple remnant-movement-avoiding derivations, or constraint rankings.
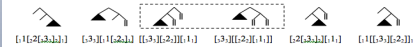
Structures generated by Cinque's analysis: Note that the order 3-2-1 has two distinct possible analyses.

- Another account is that of Abels & Neeleman (2012).
  - Similar to Cinque's analysis, but permit free head-complement order, abandoning the LCA.
  - Technically they allow a superset of Cinque's derivations, and therefore greater ambiguity.
  - Ambiguity can be reduced by adding a preference for derivations with fewer movements.

- The ULTRA account is superior on these three points.
  1. 213-avoidance falls out from the architecture; such orders cannot be parsed successfully (*as single-domain information-neutral orders*).
  2. ULTRA is universal and unparameterized, parsing the word orders of all languages.
  3. The brackets generated by ULTRA (by reading Push and Pop as left and right brackets, resp.), while largely identical to those in Cinque's (2005) analysis, are unambiguous for any single-phase order. (*Note: the item being sorted is the label.*)
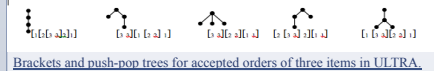
Brackets and push-pop trees for accepted orders of three items in ULTRA.

CONCLUSION

- We introduced the ULTRA model.
  - Universal parser-is-grammar without parameters or learning.
  - Stack-sorting + cartography = universal parsing.
  - Explains a class of word order universals, while eliminating spurious ambiguities.
  - The model covers linearization, displacement, composition, brackets and labels.
- Our ongoing research aims to extend the model to explain other aspects of grammar.

SELECTED REFERENCES

Abels, K. & Neeleman, A. (2012). Linear asymmetries and the LCA. *Syntax*, 15(*1*), 25-74.

Cinque, G. (2005). Deriving Greenberg's Universal 20 and its exceptions. *Linguistic Inquiry*, 36(*3*), 315-332.

Cinque, G. (2014). *Typological studies: Word order and relative clauses.* Routledge.

Greenberg, J.H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. *Universals of language*, 2, 73-113.

Knuth, Donald. (1968). The Art of Computer Programming Vol. 3: Sorting and Searching.

Steddy, S., & Samek-Lodovici, V. (2011). On the ungrammaticality of remnant movement in the derivation of Greenberg's Universal 20. *Linguistic Inquiry*, 42(3), 445-469.

Wurmbrand, S. (2006). Verb clusters, verb raising, and restructuring. *Blackwell companion to syntax*, 5, 229-343.

ACKNOWLEDGMENTS