Gradient representations in multilingual grammars

Mike Putnam, Penn State University

Heritage Language Acquisition Workshop
UiT – September 19, 2016

Overview

- Discuss the need (+ advantages and challenges) of including gradient representations in theoretical analyses of bi/multilingual grammars
- Review two previous studies (Hopp & Putnam 2015; Westergaard et al. 2016) from this perspective

Why is this important for research on heritage grammars?

- Can address the fluid nature of grammars across the lifespan
- Can reveal facilitative, non-facilitative, and emergent traits
- Amenable to experimental research (including computational work)

Where and *what* is gradient in grammars?

Not all grammatical constraints are categorical

- Well-formedness has never really been an all-or-nothing matter (*?, ??, ?*, **)
  - Magnitude estimations
  - Experimental data (ex. ERP)
  - Corpus data
- Contra Newmeyer (2003, 2005) – Non-categorical usage can (best) be explained by non-linguistic knowledge and processing efficiency

English NP-PP order

- (Data taken from Hawkins 2004 & Wasow 2002, 2009):
  - That brings Barry Bonds to the plate. (NP-PP)
  - That brings to the plate Barry Bonds. (PP-NP)
- 90% of time in English we find the NP-PP order
- This strong preference is non-categorical
- Hawkins (1994, 2004): Parsing is more efficient when shorter phrases proceed longer ones (EIC)

Chart from Wasow (2009)

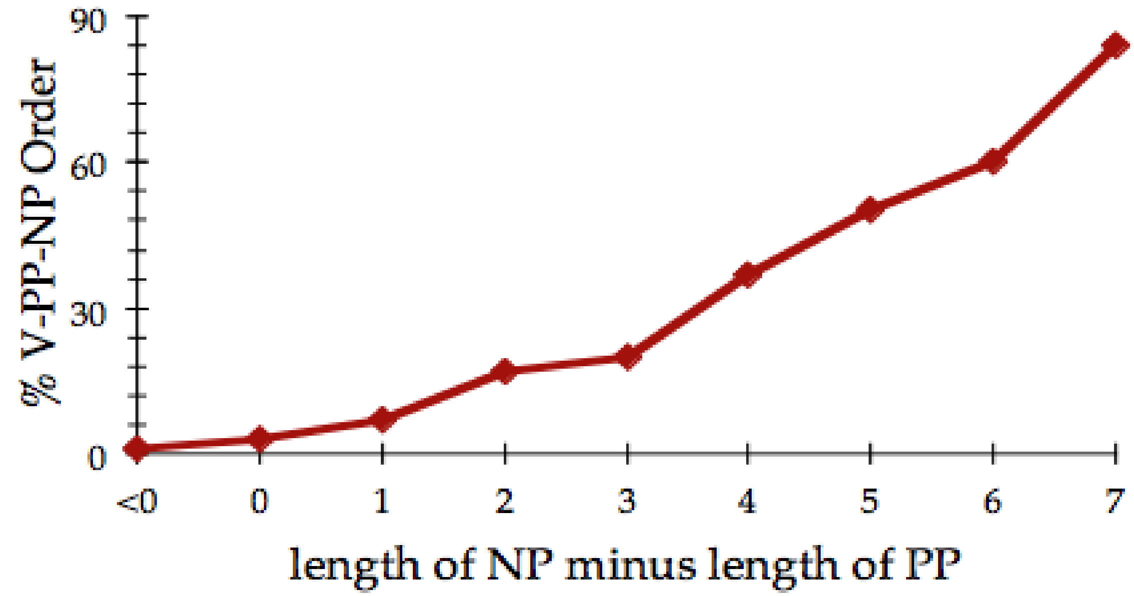

Figure 1: NP and PP Ordering in VPs from the Brown Corpus

Theoretical assumptions

Goals for linguistic theorizing

- 4 properties (Lees 1957: 376)
  - Freedom from contradiction,
  - Maximal cohesion with other branches of science,
  - Maximal validity in coverage of known data, and
  - Maximal elegance of statement

Parallel activation

- Evidence for the simultaneous, parallel activation of both/multiple languages in bi/multilinguals is pervasive (Green 1998; Dijkstra & van Heuven 2002; Blumenfeld & Marian 2007; Kroll et al. 2008; Shook & Marian 2013):
  - Phonology (Marian & Spivey 2003; Darcy et al. 2015)
  - Lexical (Linck et al. 2008; Bartolotti & Marian 2012)
  - Syntax (Koostra et al. 2012; Goldrick et al. 2016)
  - Semantic (Martin et al. 2010)

Gradient Symbolic Computation

- Integration of grammatical and gradient representations
- ICS – Integrated Connectionist/Symbolic architecture of cognition (Smolensky & Legendre 2006)
  - At the level of cognitive macro-structure, GSC incorporates not only computational but also representational principles from the micro-structure of neural-network processing.
  - Result: Blending and mixed representations

Illustration of the ICS-architecture (from SG&M 2014)

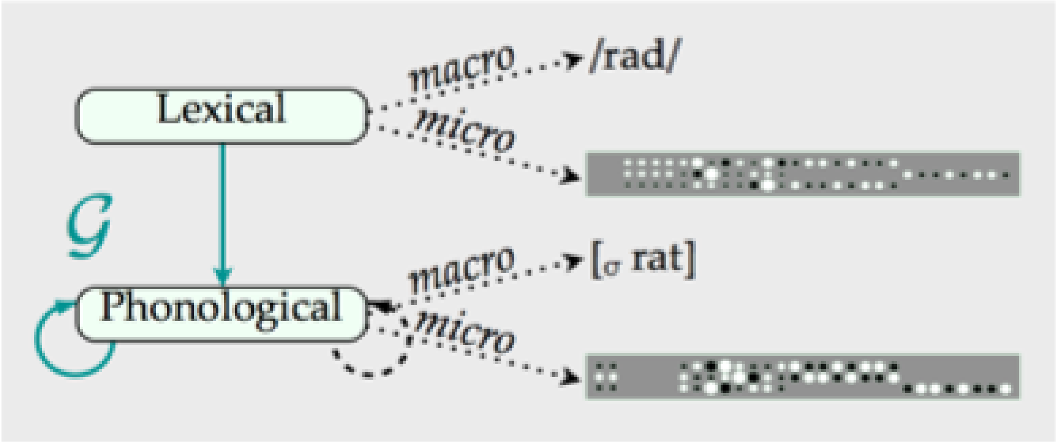

Figure 1. One parallel step of processing—one relaxation—in phonological encoding (German *Rad* 'wheel'). Input and output representations are Distributed Symbol Structures characterized at both macro- and microlevels. Evaluation (solid arrows) and quantization (dashed arrow) dynamics perform Gradient Symbol Processing.

Representations & symbols

- Q: Which elements are ideal representations and symbols?
  - It depends on your view of where competition takes place:
    - OT-type grammar
    - MP-type grammar
  - Representations: What competes?
  - Symbols: Violable constraints found in OT/HG
- Important point: A GSC-approach radically departs from a traditional OT-grammar in fundamental ways

Word order in embedded clauses in MSG (Hopp & Putnam 2015)

Overview of the H&P study

- What we were looking at?
  - Verb ordering in subordinate clauses in MSG
- Why is this interesting?
  - Matrix clause order in German is Verb-Second (V2)
  - Finite verbs appear in final position (V-last) in subordinate clauses
  - Subordinate clause word order acquired later in L1 (and L2) acquisition

H&P 2015 Spontaneous Production 1

Table 4. Frequency of word orders in subordinate clauses: Lexical verbs

Participant	Ambiguous	V2	SVO	Verb-final	Total
101				3	3
102				1	1
103	4		1	5	10
112				1	1
121	1	1	1	8	11
122	1			2	3
132				2	2
133	6	1		10	17
Totals	12	2	2	32	48

H&P 2015 Spontaneous Production 2

Table 5. Frequency of word orders in subordinate clauses: [Aux/Modal + verb]

Participant	Ambiguous	V2	SVO	Verb-final	Total
101		1			1
102	2	3		3	8
103	3	1			4
112		5		1	6
121		5		2	7
122	1	5			6
132				2	2
133	2	10		7	19
Totals	8	30	0	15	53

Summary of spontaneously produced speech

- 101 subordinate clauses
  - 67 showed ambiguous or V-last order
  - 2 instances of SVO
  - 32 cases of V2-order
- Breakdown by complementizer-type:
  - *dass* (n=17) – 15 tokens display V2-order
  - *weil* (n=9) – 8 tokens display V2-order
  - *wenn* (n=25) – 22 tokens display V-last order
  - *wo* (n=25) – 24 tokens display V-last order

GJT Experiment & Results

Table 6. Mean acceptability judgments and standard deviations. All participants (n = 8). (Note: The asterisks indicating grammaticality are based on modern standard German.)

	Mean rating	Standard Deviation
(18a) SOVAux	2.46	0.74
(18b) OVSAux	2.54	0.80
(18c) *$N_{NOM}N_{NOM}$VAux	2.58	0.99
(18d) *$N_{ACC}N_{ACC}$VAux	2.54	0.82
(19a) *SAuxAdvOV	5.08	0.77
(19b) *OAuxAdvSV	4.85	0.92
(19c) *AdvAuxSOV	5.19	0.64
(20) *SAuxVOAdv	2.94	1.46
(21) *Word salad	1.69	1.09
*S-V-Agreement	2.71	1.03

Results (con't)

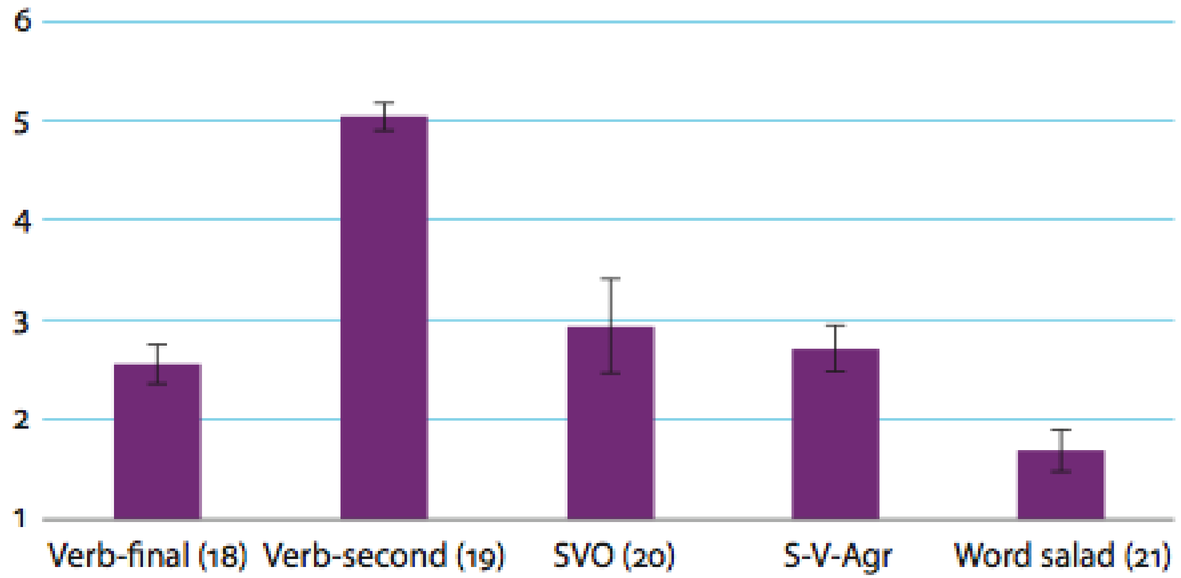

Figure 1. Mean acceptance scores for sentences in (18) through (21), all participants (n = 8, error bars show standard error of the mean)

GSC-treatment of H&P's findings

- The complementizer appears to call the shots here
- Mixed representations:
  - dass/weil S_{NP} (V * λ) … Part (V *μ)
  - wenn S_{NP} (V *λ) … Part (V * μ)
- Constraints
  - S_{NP}-V: subject before V (faithfulness)
  - Part-V: prevent part-V order (markedness)
  - Part-O_{NP}: penalize Part-O order (markedness)

Tableau 1- *dass*-clauses

Input: dass S_{NP} (V * λ) … Part (V *μ)					
German: 0.7 activation	- 5	- 3	- 25		
English: 0.3 activation	- 10	- 10	- 0		
Combined weighting	- 6.5	- 5.1	- 17.5		
	*Part-V	S_{NP}-V	*Part-O_{NP}	*H*	*P*
…dass S_{NP} V O_{NP} Part	0	- 10.2	0	- 10.2	87%
…dass S_{NP} V O_{NP} Part	0	0	- 35	- 35	1%
…dass S_{NP} Part O_{NP} V	- 13	- 10.2	0	- 23.2	12%

Multilingualism & Proximity

Westergaard et al. 2016

- The Linguistic Proximity Model (LPM)
  - Takes a closer look at the CLIs of L3A in simultaneous bilinguals
- The study:
  - Grammaticality judgment task with two word ordering conditions related to verb movement (V2 and subj-aux inversion in English)
- Participants:
  - 3 groups of 11-14 year olds
    - Norwegian-Russian bilinguals (*n*=22)
    - Norwegian monolinguals (*n*=46)
    - Russian monolinguals (*n*=31)

The (non-)facilitative role of V2

ENG = RUS ≠ NOR

Emma		**often**	eats	sweets.	ENG
Emma		**často**	jest	konfety.	RUS
Emma	spiser	**ofte**		konfekt.	NOR

Predictions

- V2 ordering w.r.t. adverbials
  - Monolingual Russians (L1Rus) should perform at ceiling (due to word order similarities between L1 and L2)
  - L1Nor should transfer the V2 property (i.e., verb movement)
  - The bilinguals (2L1) are predicted to outperform the L1Nor-participants (due to the presence of Russian).
  - 2L1s may perform worse than L1Rus (due to Norwegian influence)

Results

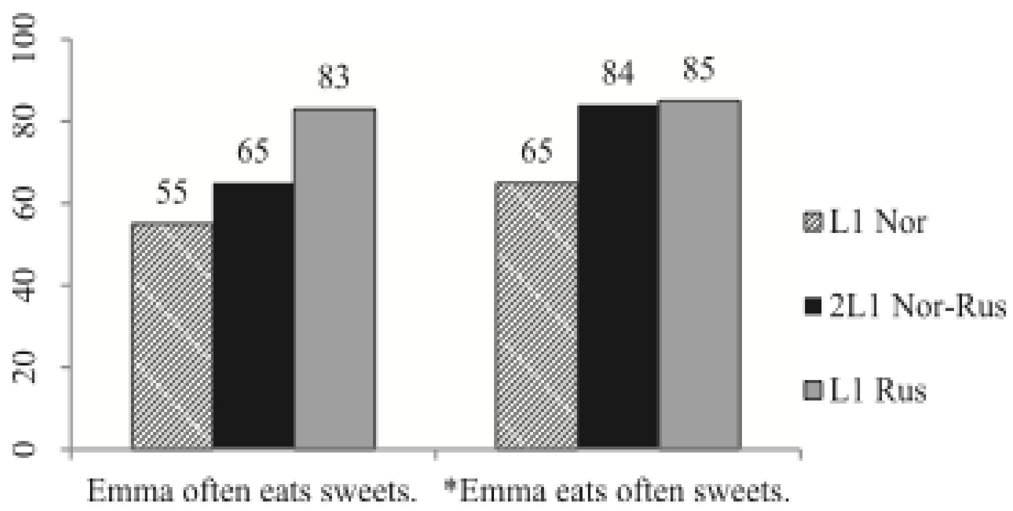

Figure 2. Percentage of correct responses in grammatical (Adv-V) and ungrammatical (V-Adv) sentences.

GSC-analysis of these findings

- Competing mixed representation:
  - S_{NP} (V *λ) Adv (V *μ) O_{NP}
- Symbols (in the form of violable constraints) evaluate the gradient representations generated from the activation of multiple grammars
- Constraints:
  - V-Adv (markedness)
  - ParseEngl (faithfulness); Adv-V

LPM & gang-up effects

- L1Nor kids
  - Norwegian 0.7 activation
  - English 0.3 activation
- 2L1 kids
  - Norwegian 0.35 activation
  - Russian 0.35 activation
  - English 0.3 activation
- Given that Russian and English share ordering, the activation values will lead to facilitating effects

Summary

- L1N kids over-accept ungrammatical English stimuli that contain V2-structures (equivalent to Norwegian; V-Adv)
- 2L1N-Rs are more successful in noticing these errors (due to the facilitating effect of Russian)
- A GSC-analysis thus subsumes the LPM due to the multiple activation of all three grammars in the trilingual population.

General discussion

- The GSC-architecture shows promise for investigations involving bi/multilingual grammars
- 4 properties (Lees 1957: 376)
  - Freedom from contradiction,
  - Maximal cohesion with other branches of science,
  - Maximal validity in coverage of known data, and
  - Maximal elegance of statement

The road ahead

- Challenges remain:
  - Although competing neural activation and activation spreading is pervasive, which method is best to represent and calculate this?
  - Re: representations – Which structures participate in these analyses (e.g., exo-cues, parallel levels, etc.)?
  - Re: symbols – What sorts of well formedness conditions are placed on the constraints that evaluate these gradient structures? Where and when are they active (and when not)?
  - Future studies need to move beyond linearization properties (Schwarz, in prep.).

Thanks!

- Special thanks to:
  - Matt Carlson
  - Matt Goldrick
  - Lara Schwarz
  - Paul Smolensky
  - Géraldine Legendre
  - LCC @ PSU lab group