Second-Language Speakers and Morphological Irregularity: A Computational Approach

Weldon Scott

AP Research

April 26, 2021

Compositionality and Irregularity

- Manner of inflection (Szabó, 2004)
- Frequency (Wu et al., 2019)

Compositional	Irregular
walk-ed	had
play-ed	was/were

Adaptability of Language

- Lupyan and Dale (2010)
- Fitting social/communicative niche
- Population size

Dinka

- 2 million native speakers (Ladd et al., 2009)
- Limited L2 acquisition

Second-Language Acquisition

- Birdsong and Flege (2001)
- Difficulty with irregular forms
- Retention of familiar features

How does the presence of second-language speakers affect the degree of irregularity in a language?

Initial Expectations

- Shaped by Lupyan and Dale
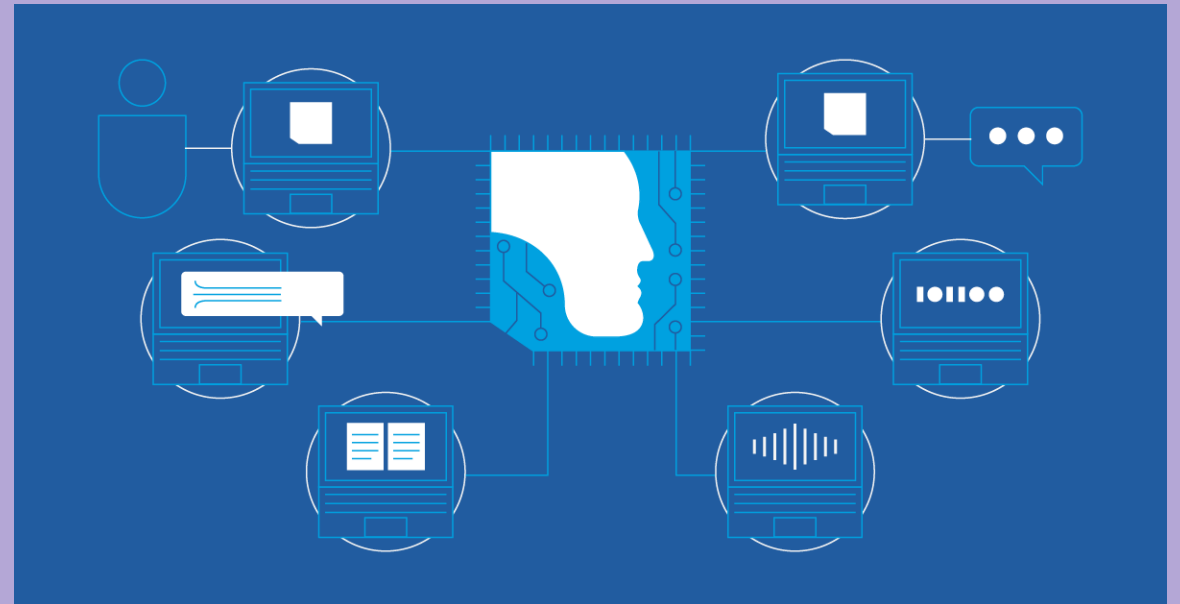
- Native language influence

- Easier learning

Natural Language Data

- Desired
- External variables (Brezina and Pallotti, 2016)
- Lack of uniformity (Clyne, 1992)

Simulated Language Data

- Quantity of data (Mannila et al., 2013)

- Versatility (Cangelosi & Parisi, 2002)

- Increasing prevalence (Mitkov, 2005)

Iterated Learning Model

- Simplistic but generalizable (Kirby, 2001)
- Iterative evolution of linguistic structure

Meaning Space

- 2-dimensional (Kirby, 2001)
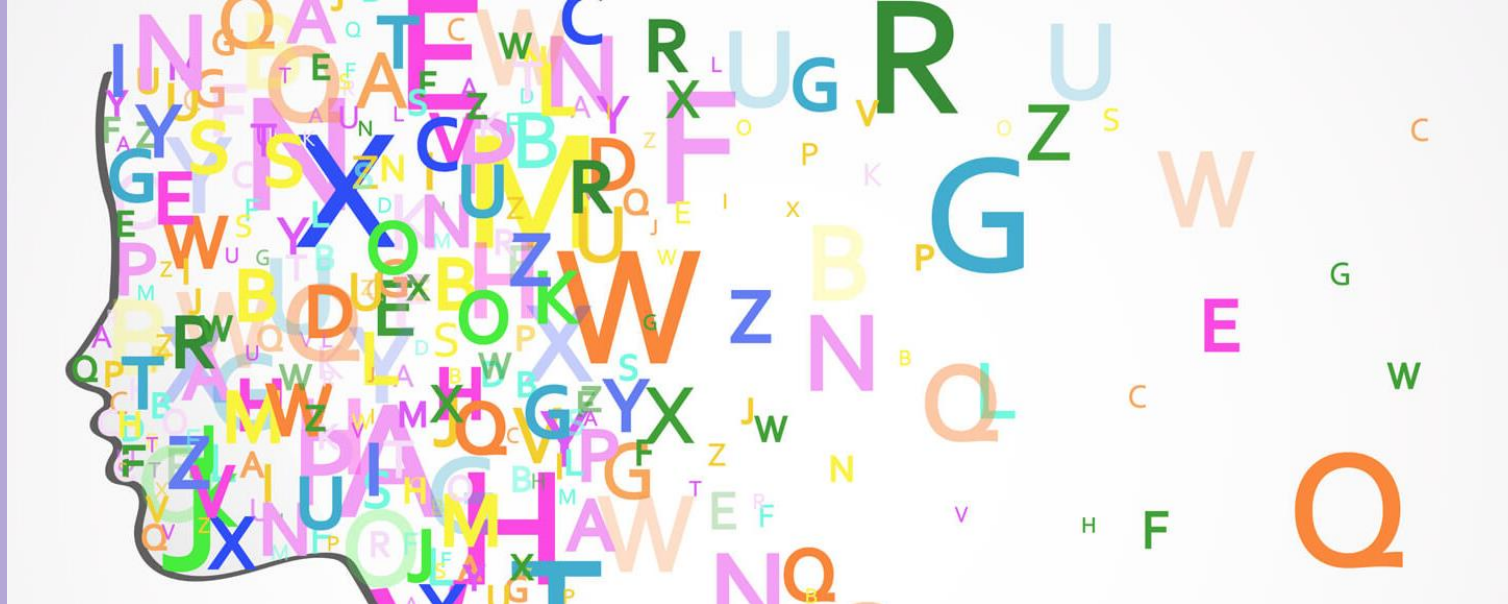- (A_0, B_0) to (A_9, B_9)
- Thoughts

Signal Space

- Sequences of characters in Latin alphabet
- "Atomic units of language" (Kirby, 2001)
- Speech

Definite Clause Grammar

- Linguistic analysis capabilities (Pereira & Warren, 1980)
- Set of rewrite rules
- Specific to individual speakers

Sample Grammar

(X, Y) → XY	
(A0, B0) → v	
A0 → s	
A1 → de	
B0 → a	
B1 → q	

	A0	A1
B0	v	dea
B1	sq	deq

Generation Structure

- Speaker and listener
- Speaker error
- Uneven distribution for meanings (Piantadosi, 2014)

Meaning Distribution

	A0	A1	A2	A3	A4	A5	A6	A7	A8	A9
B0										
B1										
B2										
B3										
B4										
B5										
B6										
B7										
B8										
B9										
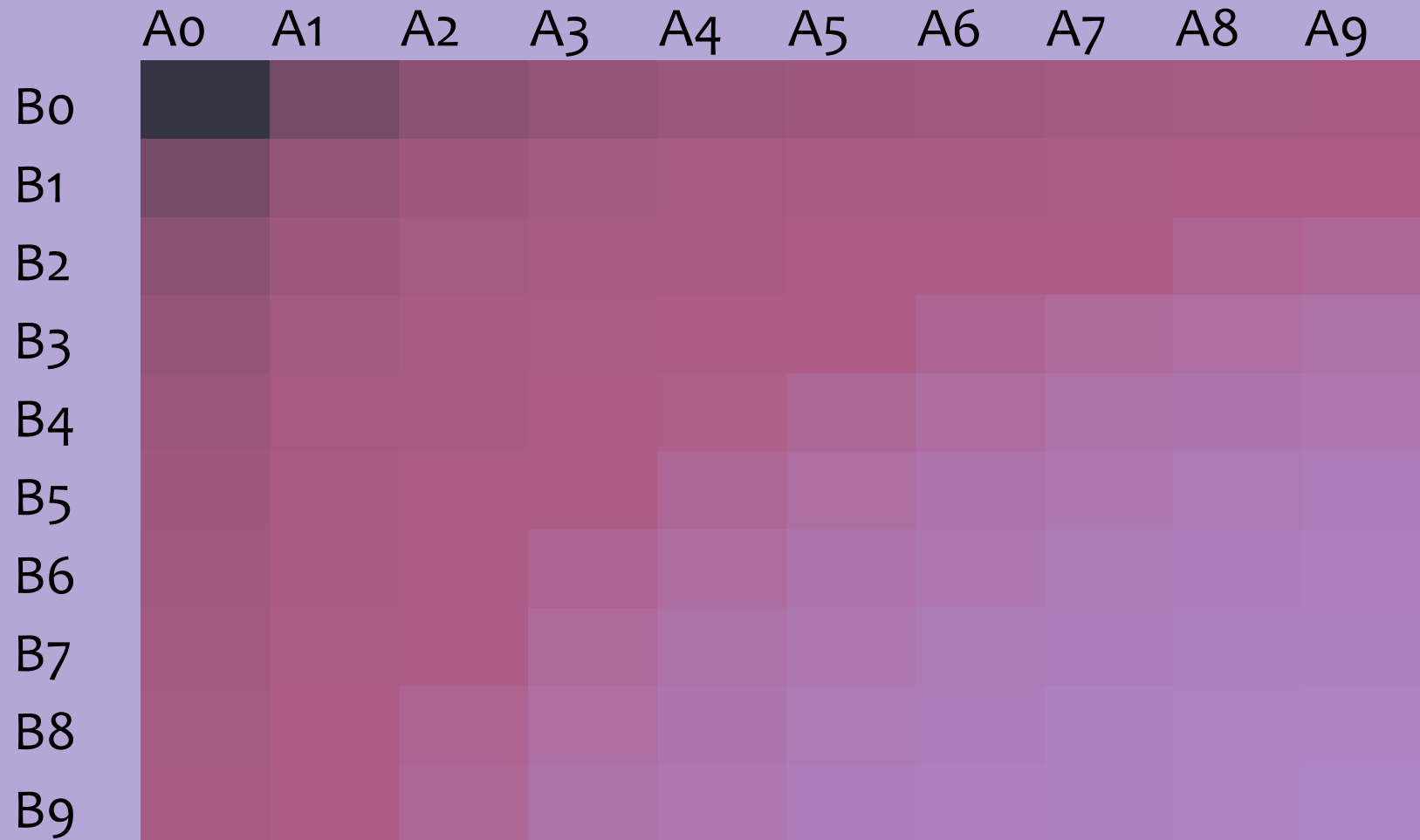
(A0, B0)	11.6%
(A1, B0)	5.8%
(A5, B0)	1.9%
(A5, B5)	0.32%
(A9, B9)	0.11%

Simulation Structure

- Homogeneous and heterogeneous stages
- 100 samples
- Signals for all meanings

Homogeneous Stage

- Initial evolution
- 10 languages
- 90% after 25 generations

Heterogeneous Stage

- Interlinguistic contact
- Most and least irregular languages
- 5% increments

Sample Language

	A0	A1	A2	A3	A4	A5	A6	A7	A8	A9
B0	t	yc	d	l	c	xc	pc	zc	jcp	qc
B1	tb	y	jb	lb	bb	xb	pb	zb	jbp	qb
B2	u	yu	ju	lu	bu	xu	pu	zu	jup	qu
B3	ty	yy	jy	ly	by	xy	py	zy	jyp	qy
B4	tm	ym	jm	lm	bm	xm	pm	zm	jmp	qm
B5	th	yh	jh	lh	bh	xh	ph	zh	jhp	qh
B6	td	yd	jd	ld	bd	xd	pd	zd	jdp	qd
B7	n	yn	jn	ln	bn	xn	pn	zn	jnp	qn
B8	tz	yz	jz	lz	bz	xz	pz	zz	jzp	qz
B9	tf	yf	jf	lf	bf	xf	pf	zf	jfp	qf

Prevalence of Irregularity

Homogeneous Stage Results

Metric	Value
Mean	6.44
Standard Deviation	1.45
Median	7
Minimum	2
Maximum	11
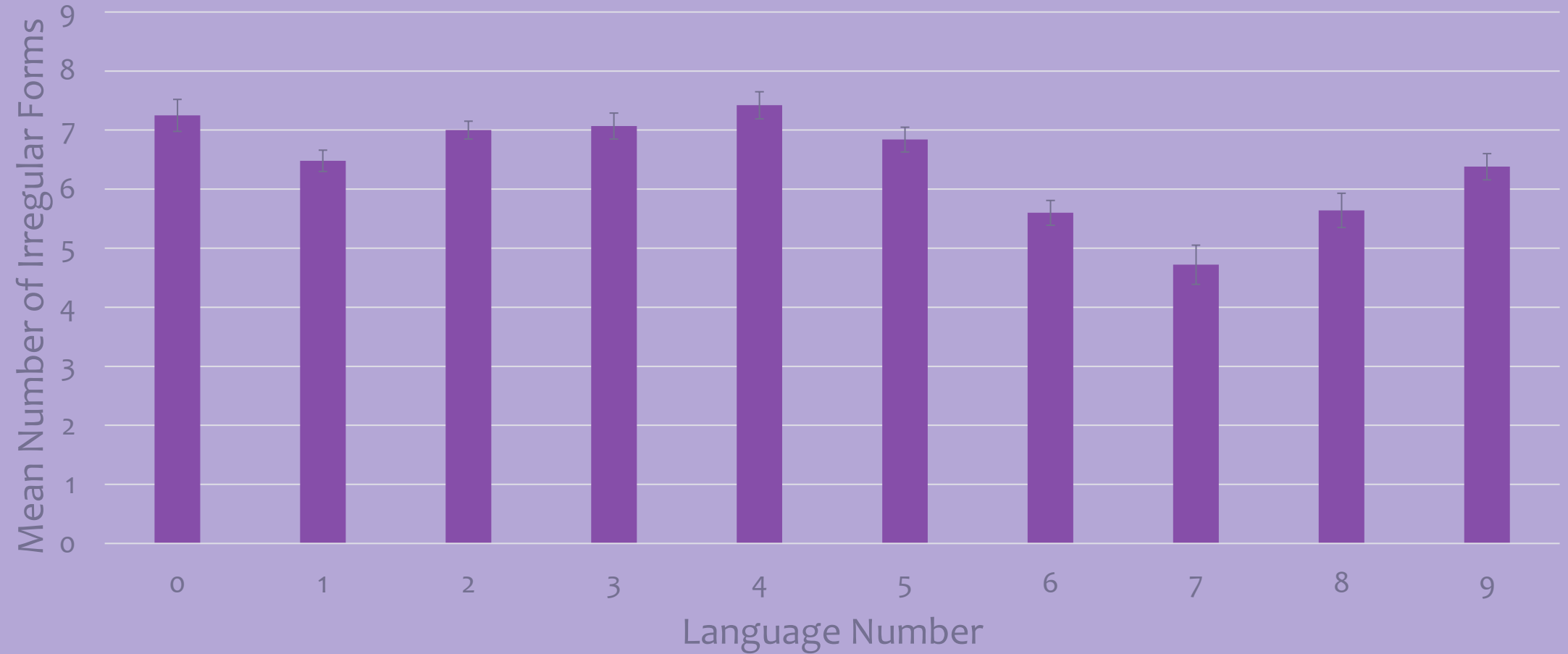
Heterogeneous Stage Selection

Metric	Language 4	Language 7
Mean (95% confidence)	7.42 ± 0.23	4.72 ± 0.33
Standard Deviation	1.16	1.70
Median	7	4

Acquisition of High-Irregularity Language
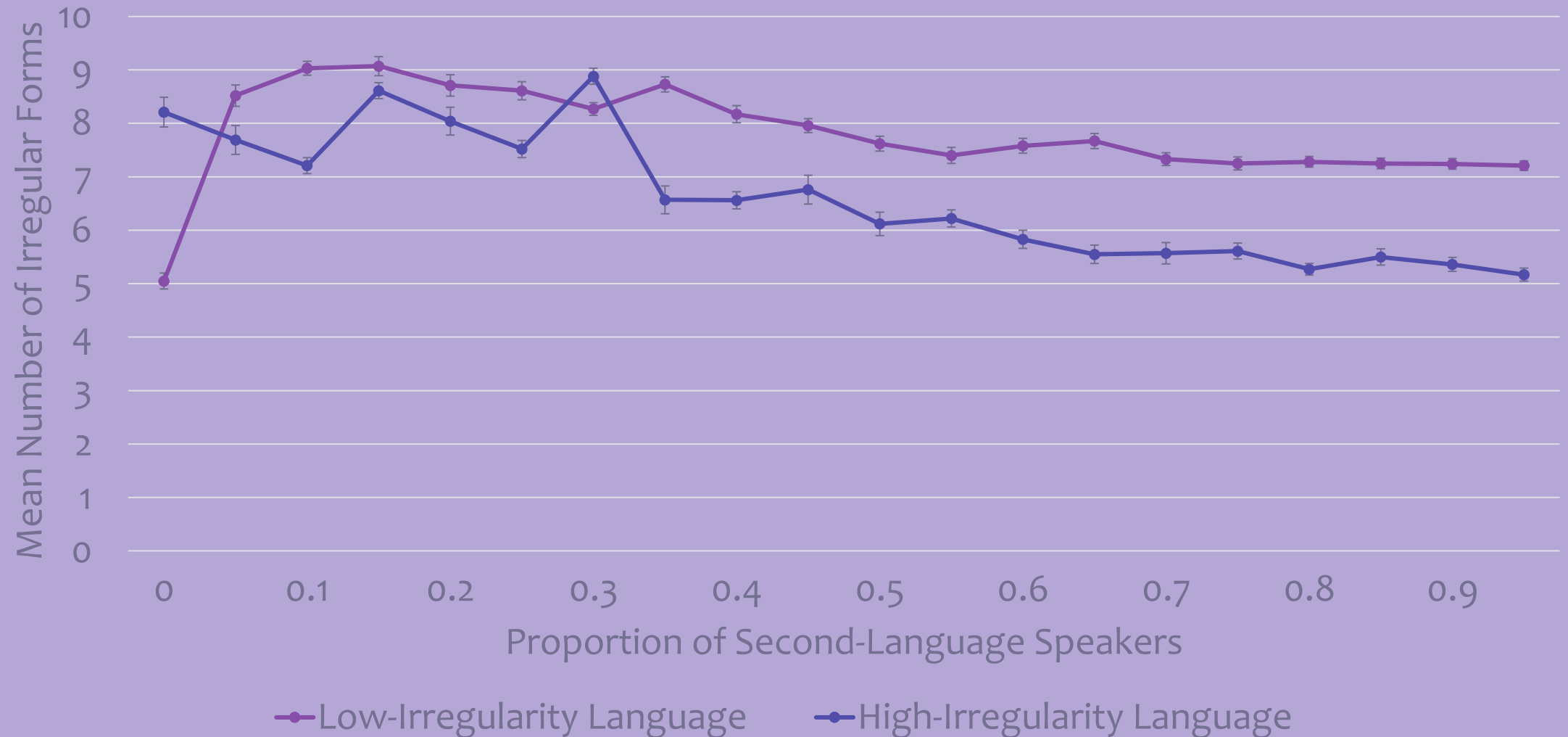
Mean Number of Irregular Forms

Proportion of Low-Irregularity Language Speakers
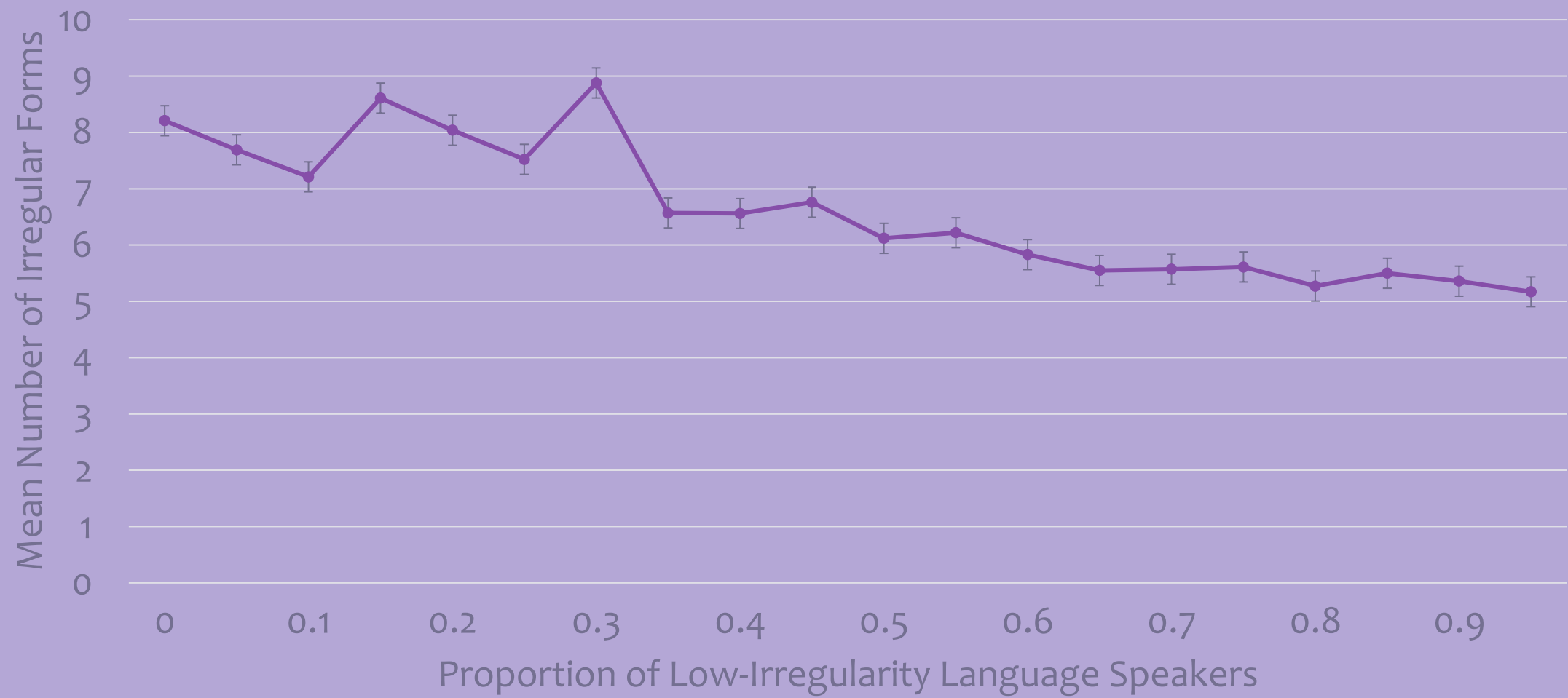
Heterogeneous Stage Results

- Unexpected increases
- Length of irregular forms
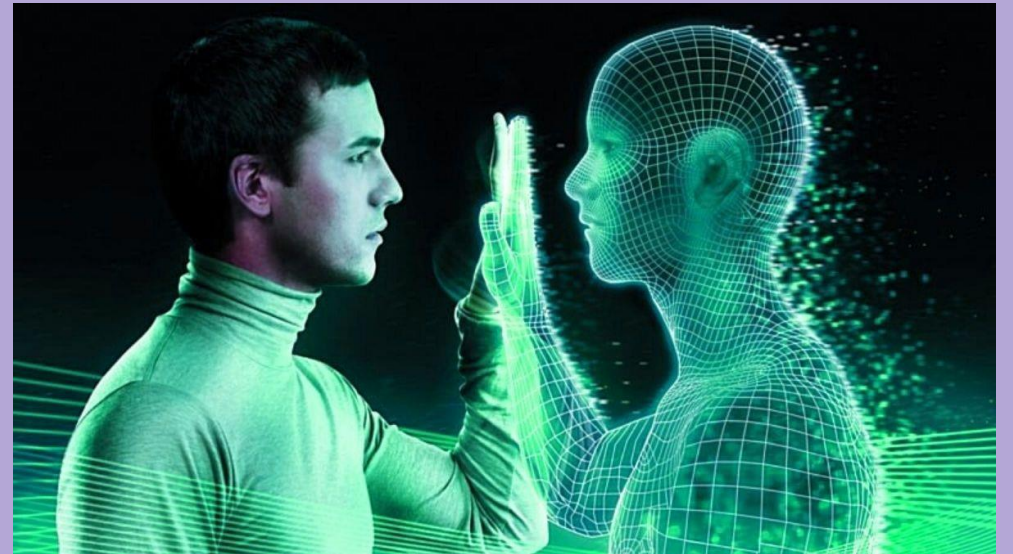
Heterogeneous Stage Results

Implications

- Analyze variables in isolation
- Support Lupyan and Dale
- Language learners

Limitations

- Arbitrary parameters (Cangelosi & Parisi, 2002)
- Small meaning space (Kirby, 2001)
- Only simple combinations of languages

Future Directions

- Complex computational models (e.g., Kirby, 2002)
- Context-dependent grammar
- Cross-linguistic studies

References

Birdsong, D., & Flege, J. E. (2001). Regular-irregular dissociations in L2 acquisition of English morphology. *BUCLD 25: Proceedings of the 25th Annual Boston University Conference on Language Development* (pp. 123–132). Boston, MA: Cascadilla Press.

Brezina, V., & Pallotti, G. (2016). Morphological complexity in written L2 texts. *Second Language Research, 35*(1), 99–119. doi: 10.1177/0267658316643125

Cangelosi A., & Parisi D. (2002) Computer simulation: A new scientific approach to the study of language evolution. In Cangelosi A., Parisi D. (Eds.), *Simulating the evolution of language*. Springer. doi: 10.1007/978-1-4471-0663-0_1

Clyne, M. G. (Ed.). (1992). *Pluricentric languages: Differing norms in different nations*. Walter de Gruyter.

Crystal, D. (2006). English worldwide. In D. Denison & R. M. Hogg (Eds.). *A history of the English language*. pp. 420–439. Cambridge University Press. doi: 10.1017/CBO9780511791154.002

Kirby, S. (2001). Spontaneous evolution of linguistic structure-an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation, 5*(2), 102–110. doi: 10.1109/4235.918430

References

Kirby, S. (2002). Learning, bottlenecks and the evolution of recursive syntax. In T. Briscoe (Ed.). *Linguistic evolution through language acquisition.* pp. 173–203. Cambridge University Press. doi: 10.1017/CBO9780511486524.006

Ladd, D. R., Remijsen, B., & Manyang, C. A. (2009). On the distinction between regular and irregular inflectional morphology: Evidence from Dinka. *Language, 85*(3), 659–670.

Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PloS One, 5*(1), e8559. doi:10.1371/journal.pone.0008559

Mannila, H., Nevalainen, T., & Raumolin-Brunberg, H. (2013). Quantifying variation and estimating the effects of sample size on the frequencies of linguistic variables. In M. Krug & J. Schlüter (Eds.), Research methods in language variation and change (pp. 337–360). Cambridge University Press.

Mitkov, R. (Ed.). (2005). *The Oxford handbook of computational linguistics.* Oxford University Press. doi: 10.1093/oxfordhb/9780199276349.001.0001

Pereira, F. C. N., & Warren, D. H. D. (1980). Definite clause grammars for language analysis—A survey of the formalism and a comparison with augmented transition networks. *Artificial Intelligence, 13*(3), 231–278. doi: 10.1016/0004-3702(80)90003-X

References

Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review, 21*(5), 1112–1130. doi: 10.3758/s13423-014-0585-6

Seidlhofer, B. (2001). Closing a conceptual gap: The case for a description of English as a lingua franca. *International Journal of Applied Linguistics, 11*(2), 133–158. doi: 10.1111/1473-4192.00011

Szabó, Z. G. (2020). Compositionality. *The Stanford encyclopedia of philosophy*, E. N. Zalta (Ed.).

Wu, S., Cotterell, R., and O'Donnell, T.J. (2019). Morphological irregularity correlates with frequency. *In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 5117–5126. Florence, Italy

Thank You