



Morphological productivity

What do paradigm gaps tell us?

Muhammed İleri (*ileri@sas.upenn.edu*)

University of Pennsylvania

Ling 203

University of Delaware

September 29, 2025

Productivity of human language

Human language is a discrete combinatorial system.

Discrete elements are combined to create larger discrete units:
sounds > morphemes > words > phrases > sentences.

Language acquisition is not solely memorization or imitation: we are able to understand and produce novel utterances.

- (1) The Martians attended the guest lecture in Newark.

Morphological productivity: Wug test (Berko 1958)

1.



This is a WUG



Now there is another one.
There are two of them.
There are two _____.

5



This is a man who knows how to RICK.
He is Ricking. He did the same thing
yesterday. What did he do yesterday?
Yesterday he _____.

Paradigm Cell Filling Problem (Ackerman et al. 2009)

Table 1. Inflection of Turkish *el* 'hand'

	UnPoss	1sgPoss	2sgPoss	3sgPoss	1plPoss	2plPoss	3plPoss	
Sg	Nom	<i>el</i>	<i>elim</i>	<i>elin</i>	<i>eli</i>	<i>elimiz</i>	<i>eliniz</i>	<i>elleri</i>
	Acc	<i>eli</i>	<i>elimi</i>	<i>elini</i>	<i>elini</i>	<i>elimizi</i>	<i>elinizi</i>	<i>ellerini</i>
	Gen	<i>elin</i>	<i>elimin</i>	<i>elinin</i>	<i>elinin</i>	<i>elimizin</i>	<i>elinizin</i>	<i>ellerinin</i>
	Dat	<i>ele</i>	<i>elime</i>	<i>eline</i>	<i>eline</i>	<i>elimize</i>	<i>elinize</i>	<i>ellerine</i>
	Loc	<i>elde</i>	<i>elimde</i>	<i>elinde</i>	<i>elinde</i>	<i>elimizde</i>	<i>elinizde</i>	<i>ellerinde</i>
Pl	Abl	<i>elden</i>	<i>elimden</i>	<i>elinden</i>	<i>elinden</i>	<i>elimizden</i>	<i>elinizden</i>	<i>ellerinden</i>
	Nom	<i>eller</i>	<i>ellerim</i>	<i>ellerin</i>	<i>elleri</i>	<i>ellerimiz</i>	<i>elleriniz</i>	<i>elleri</i>
	Acc	<i>elleri</i>	<i>ellerimi</i>	<i>ellerini</i>	<i>ellerini</i>	<i>ellerimizi</i>	<i>ellerinizi</i>	<i>ellerini</i>
	Gen	<i>ellerin</i>	<i>ellerimin</i>	<i>ellerinin</i>	<i>ellerinin</i>	<i>ellerimizin</i>	<i>ellerinizin</i>	<i>ellerinin</i>
	Dat	<i>ellere</i>	<i>ellerime</i>	<i>ellerine</i>	<i>ellerine</i>	<i>ellerimize</i>	<i>ellerinize</i>	<i>ellerine</i>
Abl	Loc	<i>ellerde</i>	<i>ellerimde</i>	<i>ellerinde</i>	<i>ellerinde</i>	<i>ellerimizde</i>	<i>ellerinizde</i>	<i>ellerinde</i>
	Abl	<i>ellerden</i>	<i>ellerimden</i>	<i>ellerinden</i>	<i>ellerinden</i>	<i>ellerimizden</i>	<i>ellerinizden</i>	<i>ellerinden</i>

Figure 1: Nominal Paradigm: distinct forms of *el* 'hand' in Turkish (Plank 1991)

84 distinct forms: we can't expect to have encountered them all during acquisition.

Based on partial input, we make generalizations (Berko 1958).

In this case, knowing 10 suffixes (case, number etc.) and how to combine them suffices.

When productivity fails: Paradigm gaps

Some forms that are expected to exist are missing with no obvious grammatical reason (Sims 2015).

The past tense of *forgo* or past participle of *stride*? (Embick & Marantz 2008, Yang 2016)

When productivity fails: Paradigm gaps

Some forms that are expected to exist are missing with no obvious grammatical reason (Sims 2015).

The past tense of 'to forgo' or past participle of 'to stride'? (Embick & Marantz 2008, Yang 2016)

спросить 'to ask'	SINGULAR	PLURAL	победить 'to be victorious'	SINGULAR	PLURAL
1 ST PERSON	спрошу / sprošu	спросим / sprosim	1 ST PERSON	*	победим / pobedim
2 ND PERSON	спросишь / sprosiš'	спросите / sprosite	2 ND PERSON	победишь / pobediš'	победите / pobedite
3 RD PERSON	спросит / sprosit	спросят / sprosjat	3 RD PERSON	победит / pobedit	победят / pobedjat

Table 1: An example of a paradigmatic gap in the Russian non-past verbal paradigm

Figure 2: A paradigm gap in Russian (Sims 2006)

Ricked vs *stridden

How do speakers know *ricked* is fine but **stridden* is not?

Negative evidence: Feedback about ungrammatical forms?

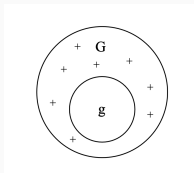


Figure 3: The Subset Problem (Pinker 1989, Yang 2016)

Ricked vs *stridden

How do speakers know *ricked* is fine but **stridden* is not?

Negative evidence: Feedback about ungrammatical forms?

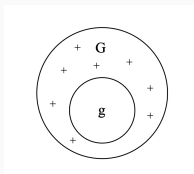


Figure 4: The Subset Problem (Pinker 1989, Yang 2016)

Kids do not receive negative feedback (Braine 1971), or don't understand/ignore them even if they do (McNeill 1966, p.69).

Child: Nobody don't like me.

Mother: No, say "nobody likes me."

Child: Nobody don't like me.

[Eight repetitions of this dialogue follow.]

Mother: No, now listen carefully, say "NOBODY LIKES ME."

Child: Oh! Nobody don't likes me.

Where does morphological productivity fail?

Unbounded creativity of word formation vs gaps.

Case study I: Spanish present indicative

There are 3 main conjugation classes in Spanish present indicatives.

Class 1 [a]

hablar 'to speak'

1sg *hablo*

2sg *hablas*

3sg *habla*

1pl *hablamos*

2pl *habláis*

3pl *hablan*

Class 2 [e]

comer 'to eat'

1sg *como*

2sg *comes*

3sg *come*

1pl *comemos*

2pl *coméis*

3pl *comen*

Class 3 [i]

vivir 'to live'

1sg *vivo*

2sg *vives*

3sg *vive*

1pl *vivimos*

2pl *vivís*

3pl *viven*

(Albright 2003)

Idiosyncratic alternations under stress

The stressed vowels exhibit alternation for some stems.

Diphthongization

$e \rightarrow ie$ [je]

sentir 'to feel'

1sg *sjento*

1pl *sentimos*

2sg *sjentes*

2pl *sentís*

3sg *sjente*

3pl *sjenten*

Diphthongization

$o \rightarrow ue$ [we]

contar 'to count'

1sg *cwento*

1pl *contamos*

2sg *cwentas*

2pl *contáis*

3sg *cwenta*

3pl *cwentan*

Raising

$e \rightarrow i$

pedir 'to request'

1sg *pido*

1pl *pedimos*

2sg *pides*

2pl *pedís*

3sg *pide*

3pl *piden*

(Albright 2003)

[rentámos] ~ [rénto] 'we/I rent'

[bendámos] ~ [béndo] 'we/I sell'

[podámos] ~ [pódo] 'we/I prune'

[montámos] ~ [mónto] 'we/I mount'

Gaps track alternating cells in 3rd-conjugation

Some 3rd-conj. verbs lack stressed forms: *abolir* 'abolish', *agredir* 'assault', *arrecirse* 'stiffen', *colorir* 'color'...

	1sg	2sg	3sg	1pl	2pl	3pl
<i>abolir</i>	*	*	*	<i>abolimos</i>	<i>abolís</i>	*
<i>dormir</i>	<i>duermo</i>	<i>duermes</i>	<i>duerme</i>	<i>dormimos</i>	<i>dormís</i>	<i>duermen</i>

(Yang 2016, p.149)

Counts from LEXESP (Yang 2016, p.149) show:

Conj.	No-chg	Diph	Raise	Total
1st (-a-)	855	84	–	939
2nd (-e-)	115	21	–	136
3rd (-i-)	12	13	10	35

Stem-vowel alternations (diphthongization/raising) are irregular/exceptional and must be memorized.

Gaps only in 3rd: no majority rule; **1st/2nd** have a productive/reliable *no-change* default (Albright 2003; Yang 2016).

Spanish gaps: conclusion

Yang 2016; Albright 2003

Learners generalize a pattern only when there is a majority rule that covers enough items (types).

Gaps occur where there is a high proportion of lexical exceptions, leading to a lack of a reliable pattern of word formation.

Gaps are exceptions that have not been memorized (i.e., not attested in the input).

Spanish gaps: conclusion

Yang 2016; Albright 2003

Learners generalize a pattern only when there is a majority rule that covers enough items (types).

Gaps occur where there is a high proportion of lexical exceptions, leading to a lack of a reliable pattern of word formation.

Gaps are exceptions that have not been memorized (i.e., not attested in the input).

Do gaps require lexical exceptions?

No—next: Turkish gaps that arise *without* lexical restriction.

Case study II: A lexically unrestricted gap - Turkish desideratives

Ileri & Demirok 2022

Turkish has a desiderative construction:

- (2) [Ben-im yemek yiy-**esi-m**] var.
1SG-GEN food eat-DESID-POS.1SG exist.cop
'I feel like eating.' Lit: My desire to eat exists.

Possible matrix predicates:

var 'exist', *yok* 'not.exist', *gel-* 'come', *git-* 'go away', *kaç-* 'run away', *tut-* 'hold', *ol-* 'to be'.

A gap in the 3PL cell

There is a *lexically unrestricted* gap in the 3PL cell of the desiderative paradigm.

	SG	PL
1	X-esim	X-esimiz
2	X-esin	X-esiniz
3	X-esi	

Table 1: The agreement paradigm of -esi desideratives.

Speakers do not converge on a well-formed output for 3PL desideratives in contexts where 3PL agreement is required (Goksel & Kerlake 2005).

Why is there a gap?

Desiderative vs. regular possessive paradigm

Irregularity in 3SG: expected ***yap-ası-sı** is ungrammatical, possibly due to haplology (Kornfilt 1986), setting up a parsing problem that yields two 3PL candidates.

	SG	PL
1	yiy-esi-m	yiy-esi-miz
2	yiy-esi-n	yiy-esi-niz
3	yiy-esi	*yiy-e-leri / *yiy-esi-leri

Desiderative paradigm.

	SG	PL
1	fare-m	fare-miz
2	fare-n	fare-niz
3	fare-si	fare-leri

Regular possessive paradigm.

Parse A: DESID = *-e*, 3SG = *-si*

⇒ 3PL: ***yiy-e-leri**

Parse B: DESID = *-esi*, 3SG = \emptyset

⇒ 3PL: ***yiy-esi-leri**

Most prefer B, others prefer A; however, most speakers reject even their preferred form. **Why is neither grammatical?**

Paradigm Cell Filling Problem in Turkish

	SG	PL
1	V-esi-m → 53.1%	V-esi-miz → 1.1%
2	V-esi-n → 2.2%	V-esi-niz → 2.1%
3	V-esi → 41.5%	*V-e(si)-leri → 0.07%

3PL desideratives are virtually absent in the input.

Speakers must infer 3PL from other cells in the paradigm.

~95% of tokens are 1SG vs. 3SG, and the latter is irregular.

An unresolved competition between two possible forms leads to uncertainty, and the gap.

Turkish gap: Conclusion

Irregularity in 3SG (i.e., an opaque haplology) results in two competing hypotheses for the unattested 3PL form.

Speakers cannot confirm their hypothesis due to absence of 3PL in the input.

Absence of conclusive evidence causes speaker uncertainty about the correct form of 3PL desideratives, which in turn leads to a gap.

This competition is implicit: speakers are unaware of the possibility of another form, unlike in lexically restricted gaps (e.g. *strode* vs *stridden*).

Discussion & conclusion

Gaps arise when there is *no reliable rule* and *no positive evidence* for the target cell; speakers cannot pick a default.

Attestation vs. gaps: if a form is attested, uncertainty yields *variation* (e.g., *dived* ~ *dove*), not a gap.

Spanish: lower frequency \Rightarrow greater uncertainty and lower acceptability (Albright 2003).

Turkish desideratives: higher frequency bases judged *less* acceptable in 3PL (I&D); consistent with *indirect negative evidence* / sensitivity to suspicious absences (Daland et al. 2007).

Discussion & conclusion

Gaps arise when there is *no reliable rule* and *no positive evidence* for the target cell; speakers cannot pick a default.

Attestation vs. gaps: if a form is attested, uncertainty yields *variation* (e.g., *dived* ~ *dove*), not a gap.

Spanish: lower frequency \Rightarrow greater uncertainty and lower acceptability (Albright 2003).

Turkish desideratives: higher frequency bases judged *less* acceptable in 3PL (I&D); consistent with *indirect negative evidence* / sensitivity to suspicious absences (Daland et al. 2007).

Takeaway: *Gaps* = missing positive evidence + no dominant rule. They are an *epiphenomenon of non-productivity*, not a separate mechanism (*possibly reinforced by indirect negative evidence, INE*).

Thank you!

Questions welcome.

Appendix: Evidence from corpus frequencies

491-million-token BOUN Web Corpus (Sak et.al. 2008)

Low freq \neq Low acceptability (Divjak 2008)

Instead of raw frequencies, relative frequencies of the forms in a paradigm should be used to detect an anomaly (Sims 2015).

	SG	PL
1	56.3%	1.16%
2	1.65%	1.82%
3	39.02%	0.07%

Table 2: Token freq. dist. of desideratives (397 types; 2,857 tokens). Only 2 3PL forms.

	SG	PL
1	0.2%	17%
2	5%	2%
3	62%	15%

Table 3: Frequency distribution of Russian *pobedit* 'to win' (adapted from Sims 2015:226).

Appendix: Overabundance of 3PL forms

Two possible forms: most prefer *X-esileri*, others *X-eleri*.

Crucially, most speakers reject even their preferred form.

They agree on the other forms in the paradigm.

	SG	PL
1	X-esim	X-esimiz
2	X-esin	X-esiniz
3	X-esi	*X-esileri / *X-eleri

Table 4: The agreement paradigm of -esi desideratives.

References i

- Ackerman, F., Blevins, J. P., & Malouf, R. (2009). Parts and wholes: Implicative patterns in inflectional paradigms. In J. P. Blevins J. Blevins (Eds.), *Analogy in grammar: Form and acquisition* (pp. 54–82). Oxford University Press.
- Albright, A. (2003). A quantitative study of Spanish paradigm gaps. In G. Garding & M. Tsujimura (Eds.), *Proceedings of the 22nd West Coast Conference on Formal Linguistics* (pp. 1–14). Cascadia Press.
- Berko, J. (1958). The child's learning of English morphology. *Word*, 14(2–3), 150–177.
<https://doi.org/10.1080/00437956.1958.11659661>
- Braine, M. D. S. (1971). On two types of models of the internalization of grammars. In D. I. Slobin (Ed.), *The ontogenesis of grammar: A theoretical symposium* (pp. 153–186). New York, NY: Academic Press.
- Daland, R., Sims, A., & Pierrehumbert, J. (2007). Much ado about nothing: A social network model of Russian paradigmatic gaps. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics* (pp. 936–943). Association for Computational Linguistics.

- Embick, D., Marantz, A. (2008). Architecture and blocking. *Linguistic Inquiry*, 39(1), 1–53.
<https://doi.org/10.1162/ling.2008.39.1.1>
- Göksel, A., & C. Kerslake. 2005. *Turkish: A Comprehensive Grammar*. New York: Routledge.
- İleri, M., & Demirok, Ö. (2023). A paradigm gap in Turkish. *Proceedings of the Workshop on Turkic and Languages in Contact with Turkic*, 7(1), 1–15.
<https://doi.org/10.3765/ptu.v7i1.5309>
- Kornfilt, J. (1986). The stuttering prohibition and morpheme deletion in Turkish. In A. Aksu-Koç & E. Erguvanlı Taylan (Eds.), *Proceedings of the Turkish Linguistics Conference* (pp. 59–83). Boğaziçi University Publications.
- McNeill, D. (1966). The creation of language by children. In J. Lyons & R. J. Wales (Eds.), *Psycholinguistic papers* (pp. 99–132). Edinburgh, Scotland: Edinburgh University Press.
- Plank, F. (Ed.). (1991). *Paradigms: The economy of inflection*. Mouton de Gruyter.
<https://doi.org/10.1515/9783110889109>

- Sak, H., Güngör, T., & Saraçlar, M. 2011. *Resources for Turkish morphological processing*. *Language Resources and Evaluation*, 45(2), 249–261.
<http://www.jstor.org/stable/41486039>
- Sims, A. (2015). *Inflectional defectiveness* (Cambridge Studies in Linguistics). Cambridge University Press. <https://doi.org/10.1017/CBO9781107053854>
- Yang, C. (2016). *The price of linguistic productivity: How children learn to break the rules of language*. MIT Press. <https://doi.org/10.7551/mitpress/9780262035323.001.0001>