The Role of Analogy in A Paradigm Gap in Turkish

Muhammed İleri
Boğaziçi Üniversitesi

20th International Morphology Meeting
Analogy in Inflection Workshop
Budapest, Hungary
3 Sept 2022

Overview

1. Introduction of -AsI desideratives
   - Irregularity in 3SG
   - Gap in 3PL
2. Empirical evidence
   - Corpus - Avoidance Strategies
   - Acceptability judgment experiment
3. The cause: Analogy
   - Uncertainty in base selection
   - Two competing forms for 3PL
4. Discussion
5. Future direction

1. INTRODUCTION OF -ASI DESIDERATIVES

-AsI Desideratives

- Turkish has a desiderative construction:

(1) [Ben-im yemek yap -ası -m] var.
1SG-GEN food do -DESID -POS.1SG exist.cop
'I feel like cooking.'

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

GEN-POSS agreement

(2) [Ben-im yemek yap -asɪ -m] var.
1SG-GEN food do -DESID -POS.1SG exist.cop
'I feel like cooking.'

- Its subject is GEN, just like in nominalizations:

(3) [Ben-im git -me -m] gerek.
1SG-GEN go -NMLZ -POS.1SG necessary.
'I need to go.'

- Elsewhere: Genitive Subject → Possessive Agreement

> Therefore, we expect the agreement in the desiderative construction to come from the possessive paradigm.

Agreement paradigm of desideratives

- But, there is an irregularity in the 3SG forms reminiscent of -sI deletion in Turkish (Kornfilt 1986, Göksel 2009).
- The expected form *yapasısı is ungrammatical.

	SG	PL
1	yap-ası-m	yap-ası-mız
2	yap-ası-n	yap-ası-nız
3	yap-ası	?

Table 1: The agreement paradigm of -AsI desideratives.

	SG	PL
1	yap-ma-m	yap-ma-mız
2	yap-ma-n	yap-ma-nız
3	yap-ma-sı	yap-ma-ları

Table 2: The regular possessive agreement paradigm exemplified by -mA nominalizations.

The Gap in the 3PL cell

	SG	PL
1	yap-asɪ-m	yap-asɪ-mɪz
2	yap-asɪ-n	yap-asɪ-nɪz
3	yap-asɪ	?

Table 3: The agreement paradigm of -AsI desideratives.

- When forced, speakers produce (4-a) or (4-b) for 3PL.DESID. Most judge even their preferred form unacceptable.

(4) a. ??(Onlar-ın) yemek yap -asɪ -larɪ var.
They-GEN food do -DESID -POS.3PL exist
'They feel like cooking.'

b. ??(Onlar-ın) yemek yap -a -larɪ var.
They-GEN food do -DESID -POS.3PL exist
'They feel like cooking.'

- The gap is not lexically restricted unlike most gaps in other languages (Baerman et.al. 2010, Boyé & Hofherr 2010, Halle 1973, Sims 2015).

Is it really a gap?

If a lexeme L *does not have a grammatical realization when combined with a set of morpho-syntactic and morpho-semantic features* F *that is well defined and required by syntax in a certain environment, then the paradigm cell* L+F *combination yields is defective.* (Sims, 2015)

- Desideratives co-occur with other agreement markers.
- Verbal roots such as iste- 'to want' co-occur and agree with 3PL subjects frequently.
- 3PL agreement suffix is obligatory in pro-drop contexts (Göksel & Kerslake 2005:117).

(5) *(Onlar-ın) yemek yap-ası var.
They-GEN food do-DESID.3SG exist
Intended: 'They feel like cooking.'

Thus, if 3PL.DESID not acceptable → paradigm gap.

2. EMPIRICAL EVIDENCE

Corpus Data

- 491-million-token BOUN Web Corpus (Sak et.al. 2008) on *tscorpus.com*.
- **Low freq** $\neq$ **Low acceptability** (Bader & Häussler 2009, Bermel & Knittl 2012, Divjak 2008, Manning 2003)
- 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 4: Token frequency distribution of desideratives. 397 types, 2857 tokens in total. Only 2 3PL.

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

Table 5: Frequency distribution of the verb *pobedit'* 'to win'. (adapted from Sims (2015, p.226))

Parallel constructions and avoidance strategies

	SG	PL
1	19.3%	9.6%
2	1.7%	6.0%
3	50.3%	13.1%

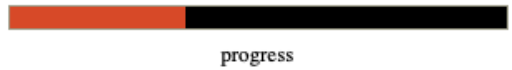
Table 6: Frequency distribution of *iste-* 'to want', another means of expressing a desire. 985,999 tokens.

	SG	PL
1	2.0%	2.2%
2	0.4%	1.8%
3	81.5%	12.1%

Table 7: Frequency distribution of NOMINATIVE -mA nominalizations. 3,447,687 tokens.

- Periphrastic constructions with *iste-*, which can bear 3PL agreement marker and convey the same meaning, is used.
- When there is an overt 3PL subject, 3SG agreement marker can be (and mostly will be) used.

Acceptability Judgment Experiment



Haftaya bisiklet alasıları varmış.

(kulağa doğal gelmiyor) 1 2 3 4 5 6 7 *(kulağa doğal geliyor)*

- 12 high freq. 12 low freq. mono-transitive verbs tested
- Sentence length, word types, word order controlled
- 48 fillers, 2 groups, randomized
- 183 uni. students aged between 18-32 (M = 21.4, SD = 2.0)
- Run online on PCIbex (Zehr & Schwarz, 2018)

(i) Haftaya bisikleti alasıları varmış. **(3PL x Frequent)**

(ii) Haftaya bisikleti alası varmış. **(Other x Frequent)**

(iii) Bugün ufaklığı şımartasıları varmış. **(3PL x Infrequent)**

(iv) Bugün ufaklığı şımartasın varmış. **(Other x Infrequent)**

Results

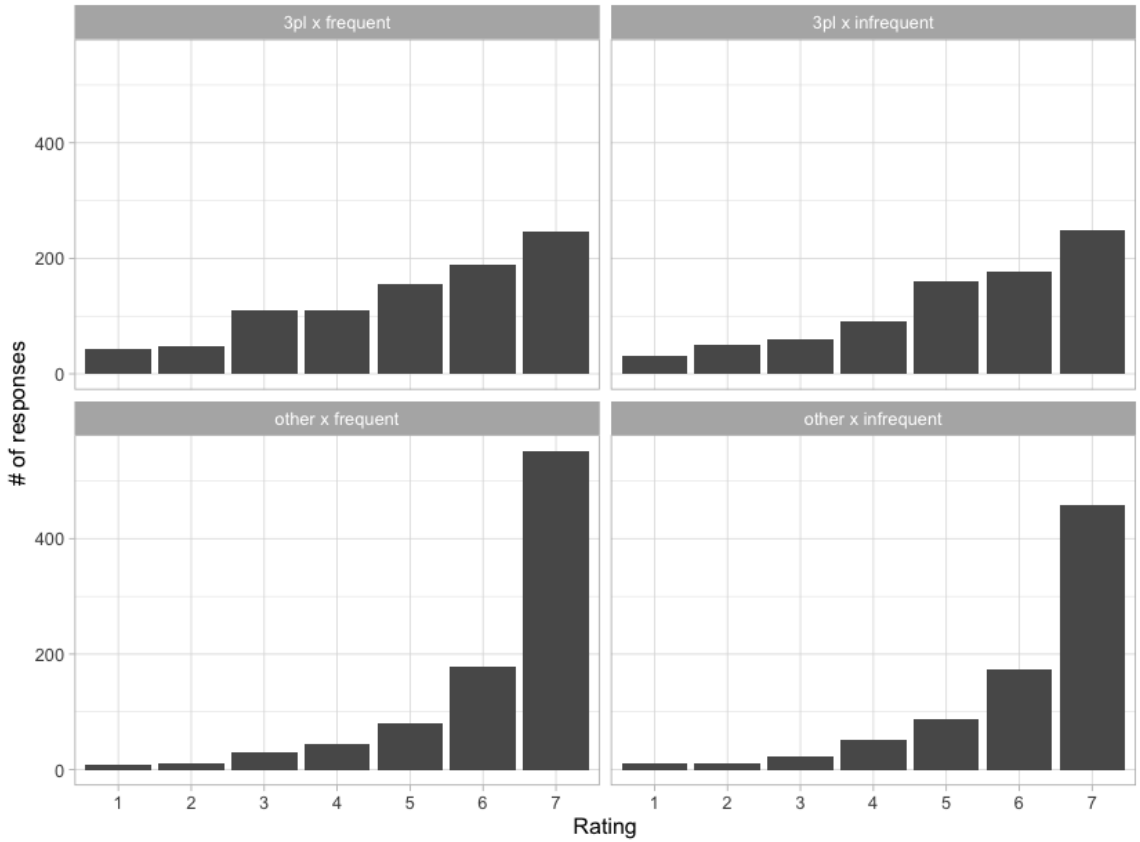


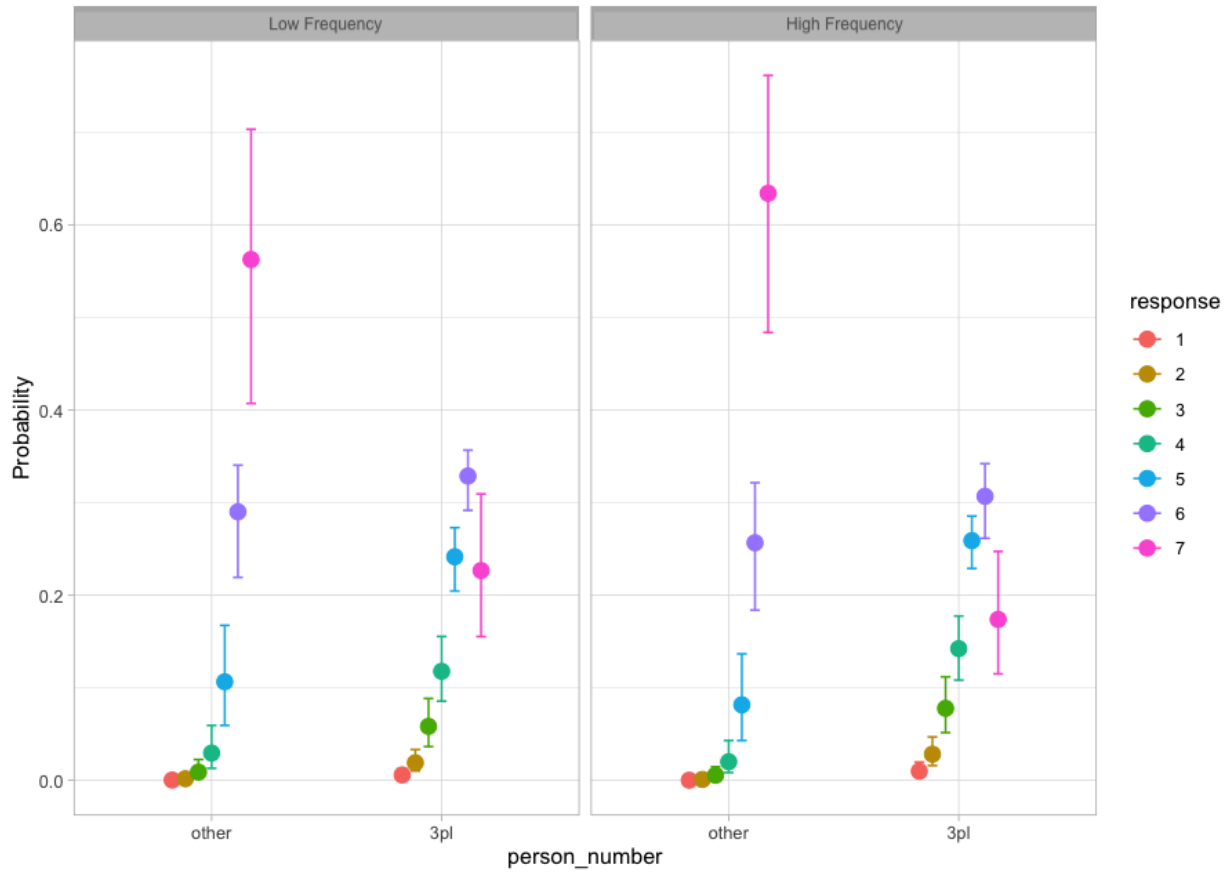
Figure 1: The distribution of raw responses per condition

- 3PL items rated lower than OTHER items.
- 6-7 ratings dominate OTHER. They are less than half in 3PL.
- No visible frequency effect.

Results

- Hierarchical Ordinal Bayesian model (Bürkner & Vuorre, 2019)

$$\text{RESP} \sim 1 + \text{AGR}*\text{FREQ} + (1 + \text{AGR}*\text{FREQ} | \text{SUBJ}) + (1 + \text{AGR} | \text{ITEM})$$



- 3PL: mean = -0.91; 95%-CI: [-1.22, -0.61]
- FREQUENT: mean=0.24 95%-CI: [-0.32, 0.65].
- FREQUENT:3PL: mean = -0.34; 90%-CI = [-0.66, -0.03].

3. THE CAUSE: ANALOGY

Variation

- There are two attested forms for 3PL desideratives: **V-ası-ları** and **V-a-ları**
- Most prefer V-ası-ları forms; some prefer V-a-ları.
- A majority is unwilling to accept even their preferred form.
- Variation restricted to 3PL desideratives. Speakers agree on single forms in other cells of the paradigm.
- Speakers are consistent in their preference.

What causes speaker variation and uncertainty in 3PL?

Paradigm Cell Filling Problem (Ackerman et. al. 2009)

	SG	PL
1	**V-asɪ-m → 56.3%**	V-asɪ-mɪz → 1.16%
2	V-asɪ-n → 1.65%	V-asɪ-nɪz → 1.82%
3	**V-asɪ → 39.02%**	??V-a(sɪ)-larɪ → 0.07%

Table 8: The agreement paradigm of -AsI desideratives including relative token frequencies of each cell.

- 3PL desideratives are virtually absent in the input.
- Many speakers have to rely on other forms to derive the 3PL.
- ∼95% of the forms belong to 1SG and 3SG, with a close competition between them.

Form availability reduces the number of options for base selection to two: **1SG** and **3SG**. Which one to choose?

Conflicting forms as a result of 4-part analogy

	SG	PL
1	2.0%	2.2%
2	0.4%	1.8%
3	81.5%	12.1%

Table 9: Freq. dist. of NOM -mA nominalizations (3,447,687 tokens).

- A simple 4-part analogy between desideratives and a very frequent similar construction, e.g. -mA nominalization.
- If analogy is made between the 1SG and 3PL, the predicted form of 3PL desiderative would be *yapasıları*:
  - yapmam : yapmaları :: yapasım : ? = **yapasıları**
- But, if analogy is made between 3SG and 3PL, the speaker would posit *yapaları*:
  - yapması : yapmaları :: yapası : ? = **yapaları**

Unresolved base competition → uncertainty → gap

(i) A majority prefers *yapasıları*; a minority prefers *yapaları*.

(ii) Many speakers find both forms unacceptable even though one is always better.

- Based on (i), I posit that 1SG is favored by speakers as a base in desideratives.
- Based on (i) and (ii), I propose that the close competition between bases is reflected as speaker uncertainty. "Paradigm gaps represent one extreme in a spectrum of uncertainty" (Albright 2003, pp.11).
- Also, uncertainty → avoidance → negative evidence (abnormal freq. dist.) → gap (Daland et. al. 2007).

Is this analysis compatible with the findings regarding base selection?

Factors affecting base selection

- Multiple factors interact with each other:

  i) Token frequency (Albright 2002, Mańczak 1980)

  - **1SG** based on desiderative frequency but **3SG** if the regular possessive paradigm, or -mA nominalization, is considered.

  ii) Morphosyntactic unmarkedness (Bybee & Brewer 1980, Tiersma 1982)

  - **3SG** more likely to be unmarked crosslinguistically and has zero exponence in many paradigms in Turkish.

  iii) Suffixes on the base (Bybee 1985, Hayes 1995, Mańczak 1958)

  - **3SG** is shorter, maybe has even zero exponent on *yapası*.

  iv) Informativeness (Ackerman et.al. 2009, Albright 2002)

  - There is probably haplology in 3SG desideratives. **1sg** has no morphophonological alternation; preserves more contrasts.

4. DISCUSSION

Discussion

- That most speakers prefer 1SG as base might be due to:
  - the criteria that favor 1SG, e.g. frequency, is more important in determining the base than others.
  - hearing other forms like the 1PL *yapasımız* suggests 1SG is a better predictor.
  - it is more likely to hear *yapasıları* forms in the input,
- Based on base selection criteria, there is no conclusive evidence that one form should be selected over the other.
- Absence of conclusive evidence causes speaker uncertainty about the correct form of 3PL desideratives.
- Speakers cannot confirm their hypothesis due to scarcity of 3PL in the input, which leads to avoidance and the gap.

Discussion

- A problematic case for Albright's (2002) *Single Surface Base Hypothesis*:
  - No speaker, not even *yapaları* preferring ones, makes errors such as **yapan* or **yapamız*
  - All speaker converge on the same forms in other cells that cannot be produced by the same mechanism with 3SG as the base.
    - yapması : yapmamız :: yapası : ? = ***yapamız**
  - Thus, *yapaları* speakers have at least two base forms.
- The gap poses a challenge to Item-and-Arrangement models such as Distributed Morphology (Halle & Marantz 1993) or Nanosyntax (Starke 2018, Caha 2019).

5. FUTURE DIRECTION

Future Direction

- A more controlled experiment to test frequency-gap interaction
- Investigate the same construction in other Turkic languages such as Turkmen and Tatar to understand how the gap emerged.
- Design an experiment to test the effect of frequency on base selection.
- Implement formal models of analogical rule learning such as Minimal Generalization Learner (Albright & Hayes 2002).

References

- Albright, A., & Hayes, B. (2002). *Modeling English Past Tense Intuitions with Minimal Generalization.* SIGMORPHON.
- Albright, A. & Bruce H. (2003). *Rules vs. analogy in english past tenses: a computational/experimental study.* Cognition 90. 119–161.
- Adam Albright (2003). *A Quantitative Study of Spanish Paradigm Gaps.* In WCCFL 22: Proceedings of the 22nd West Coast Conference on Formal Linguistics. Cascadilla Press.
- Bader, M., & Häussler, J. (2009). *Toward a model of grammaticality judgments.* Journal of Linguistics, 45, 1–58.
- Baerman, M. and Corbett, G. G. (2010). *Introduction: Defectiveness: Typology and diachrony.* In Baerman, M., Corbett, G. G., and Brown, D., editors, Defective Paradigms: Missing forms and what they tell us, pages 1–18. Cambridge University Press.

References

- Bermel, N., & Knittl, L. (2012). *Morphosyntactic variation and syntactic constructions in Czech nominal declension: Corpus frequency and native-speaker judgments.* Russian Linguistics, 36(1), 91–119.
- Boyé, G., & Hofherr, P.C. (2010). *Defectiveness as Stem Suppletion in French and Spanish Verbs*. In Defective Paradigms: Missing Forms and What They Tell Us. British Academy.
- Bürkner, P. & Vuorre, M. C. 2019. *Ordinal regression models in psychology: a tutorial.* Adv. Methods Pract. Psychol. Sci. 2, 77–101. https://doi.org/10.1177/ 2515245918823199.
- Bybee, J.L., & Brewer, M.A. 1980. *Explanation in morphophonemics: Changes in provençal and Spanish preterite forms.* Lingua, 52, 201-242.

References

- Bybee, Joan. 1985. *Morphology: A Study of the Relation between Meaning and Form.* Amsterdam and Philadelphia: John Benjamins.
- Caha, Pavel. 2019. *Case competition in Nanosyntax: A study of numerals in Ossetic and Russian.* Lingbuzz/004875.
- 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. 936-943.
- Divjak, D. 2008. *On (in)frequency and (un)acceptability.* In B. Lewandowska-Tomaszczyk (Ed.), Corpus linguistics, computer tools and applications—State of the art (pp. 213–233). Frankfurt: Peter Lang
- Göksel, A. (2009). *Compounds in Turkish*. Lingue e Linguaggio, 2, 213-236.

References

- Göksel, A. & Kerslake, C. (2005). *Turkish: A Comprehensive Grammar.* NY: Routledge.
- Göksel, A. 2009. *Compounds in Turkish.* In Lingue e linguaggio, Rivista semestrale. 2/2009. pp. 213-236. doi: 10.1418/30928
- Halle, M. 1973. *Prolegomena to a Theory of Word Formation*. Linguistic Inquiry, 4(1), 3–16. http://www.jstor.org/stable/4177749
- Halle, M. & Marantz, A. 1993. Distributed morphology and the pieces of inflection. In The View from Building 20, MIT Working Papers in Linguistics, (pp. 111–176).,Cambridge, MA: MIT Press.
- Hayes, B. (1995). *On what to teach the undergraduates: Some changing orthodoxies in phonological theory.* In Linguistics in the Morning Calm 3, 59–77.

References

- Kornfilt, J. (1986). *The Stuttering Prohibition and morpheme deletion in Turkish.* In Proceedings of the Turkish Linguistics Conference. A. Aksu Koç and E. Erguvanlı Taylan,(eds). Istanbul: Boğaziçi University Publications, 59–83.
- Löwenadler, J. (2010). *Relative Acceptability of Missing Adjective Forms in Swedish.* In Defective Paradigms: Missing Forms and What They Tell Us. British Academy.
- Manning, Christopher D. 2003. Probabilistic syntax. In Rens Bod, Jennifer Hay Stefanie Jannedy (eds.), Probabilistic linguistics, 289–341. Cambridge, MA: MIT Press.
- Mańczak, W. 1958. *Tendences générales des changements analogiques.* Lingua 7, 298–325 and 387–420.
- Mańczak, W. (1980). *Laws of analogy.* In J. Fisiak (Ed.), Historical morphology, pp. 283–288. The Hague: Mouton.

References

- 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: Cambridge University Press. doi:10.1017/CBO9781107053854
- Starke, Michal. 2018. *Complex Left Branches, Spellout, and Prefixes.* In Lena Baunaz, Karen De Clercq, Liliane M. V. Haegeman & Eric Lander (eds.), Exploring nanosyntax (Oxford Studies in Comparative Syntax), New York: Oxford University Press. DOI: https://doi.org/10.1093/oso/9780190876746.003.0009

References

- Tiersma, P. (1982). *Local and general markedness.* Language 58(4), 832–849.
- Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX). https://doi.org/10.17605/OSF.IO/MD832