

A Paradigm Gap in Turkish

Muhammed İleri & Ömer Demirok
Boğaziçi Üniversitesi

7th Workshop on Turkic and Languages in Contact with Turkic
University of Connecticut
18-19 Feb 2022

What is this talk about?

We will provide data about a gap in the 3PL cell of the agreement paradigm of -Asl desideratives in Turkish. We will propose that several factors contribute to the existence of the gap.

- ▶ Introduction of -Asl desideratives
- ▶ The gap
- ▶ Evidence from corpus data
- ▶ Evidence from an experiment
- ▶ Discussion & Conclusion

Desideratives

- ▶ Turkish has a desiderative construction:

(1) [Ben-im şarkı söyle -*yesi* -*m*] var.
1SG-GEN song sing -DESID -POS.1SG exist.cop
'I feel like singing'

- ▶ Possible matrix predicates:

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

Agreement Paradigm of Desideratives

- ▶ The expectation is met for 1st and 2nd persons. However, there is an irregularity in the 3SG forms, which is reminiscent of -sl deletion in Turkish (Kornfilt 1986, Göksel 2009).

- (4)
- a. Sen-in şarkı söyle **-yesi** **-n** varmış.
Sen(SG)-GEN song sing -DESID -POS.2SG exist.cop
'Presumably, you feel like singing.'
- b. *Pelin-in şarkı söyle **-yesi** **-si** varmış.
Pelin-GEN song sing -DESID -POS.3SG exist.cop
Intended: 'Presumably, Pelin feels like singing.'
- c. Pelin-in şarkı söyle **-yesi** varmış.
Pelin-GEN song sing -DESID.3SG exist.cop
'Presumably, Pelin feels like singing.'

Decomposing 3SG Desideratives

Two possible parses might be hypothesized for 3SG desideratives:

- (5) a. Pelin-in şarkı söyle -**yesi** -**∅** varmış.
Pelin-GEN song sing -DESID -POS.3SG exist.cop
- b. Pelin-in şarkı söyle -**ye** -**si** varmış.
Pelin-GEN song sing -DESID -POS.3SG exist.cop

'Presumably, Pelin feels like singing.'

What about 3PL desideratives?

3PL Desideratives

Depending on the decomposition of 3SG, two possible forms for 3PL can be hypothesized. However, neither of them is 'good'!

► If -sl in 3SG desideratives is POSS.3SG: -sl → -lArI

(6) ??(Çocuklar-ın) oyna -ya -lari yokmuş.
(Children-GEN) play -DESID -POS.3PL not.exist
Intended: 'They (children) don't feel like playing.'

► If -sl in 3SG desideratives is a part of DESID and POSS.3SG has a null allomorph: \emptyset → -lArI

(7) ??(Çocuklar-ın) oyna -yasi -lari yokmuş.
(Children-GEN) play -DESID -POS.3PL not.exist
Intended: 'They (children) don't feel like playing.'

The Gap

(i) ?? V + AsIIArI

(ii) ?? V + AIArI

- ▶ Speakers do not converge on a well-formed output for 3PL desideratives.
- ▶ Most prefer -AsIIArI over -AIArI although even the best option is judged to be weird or unacceptable.
- ▶ There is no free variation since usually there is not even a single acceptable form, let alone multiple forms.

Hence, we propose that there is a paradigm gap in the 3PL cell of the agreement paradigm of desideratives. This gap is not lexically restricted contrary to the widespread observation about the gaps in other languages (Halle 1973, Boyé & Hofherr 2010, Löwenadler 2010, Sims 2015).

Is the form motivated in the first place?

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)

- ▶ No morpho-syntactic reason: Desideratives co-occur with other person-number agreement suffixes.
- ▶ No semantic reason: 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).

According to the criteria above, 3PL desideratives are expected to have an acceptable form. However, they do not have one: A paradigm gap in Turkish.

Why is there a Gap?

- ▶ Decomposition of DESID.3SG is not definitive.
- ▶ Desiderative constructions are not very frequent.
- ▶ Overt 3PL agreement is only necessary in pro-drop environments.

We argue that a combination of these factors contribute to speaker uncertainty about the correct form of 3PL desideratives, which encourage speakers to avoid using the form. In turn, avoidance causes a decrease in usage, which further contributes to the persistence of the gap (Daland et.al. 2007).

Corpus Data

- ▶ We searched corpora to find evidence for the anomalous behavior of the 3_{PL} cell of -Asl desideratives.
- ▶ BUT: **Low freq** \neq **Low acceptability** (Divjak 2008, Bader & Häussler 2009, Bermel & Knittl 2012)
- ▶ Instead of raw frequencies, relative frequencies of the forms in a paradigm should be used to detect an anomaly (Sims 2015).
- ▶ One caveat: The low frequency and ambiguity of -Asl (-A [OPT] + -sl [POSS.3SG]) cause inaccurate parsings, resulting in low accuracy for our purposes.

We decided to search the largest corpus available (3.3 billion tokens) although it is not morphologically annotated: TrTen-Ten (via SketchEngine.eu)

Frequency Distribution

- ▶ Only 17 tokens for -AsIIArI in total.

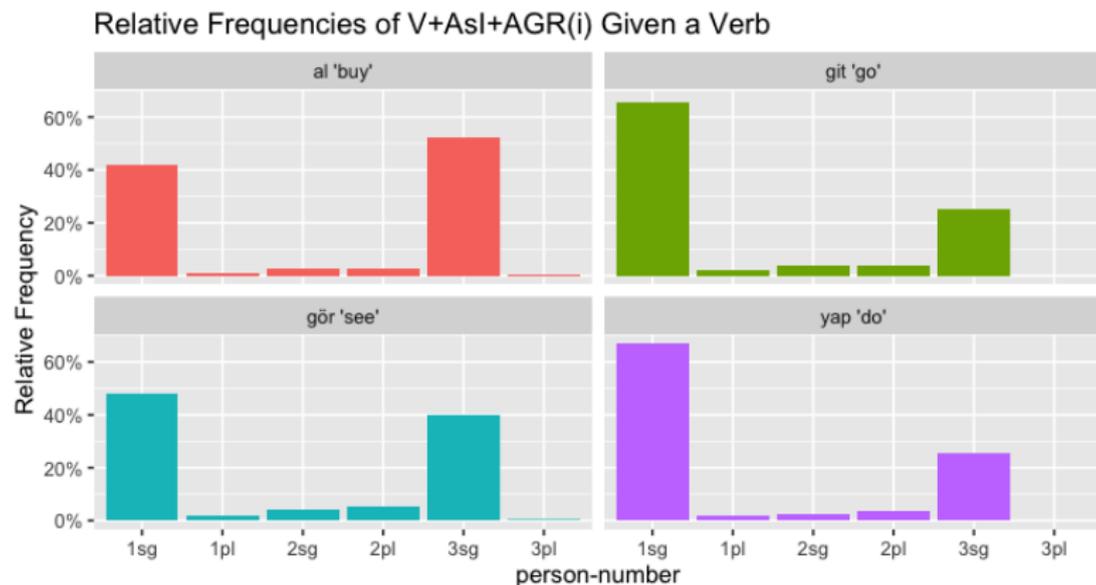


Figure: Relative frequencies of desiderative forms of 4 highly frequent verbs (Aksan et.al. 2017): *al-* 'buy', *git-* 'go', *gör-* 'see', *yap-* 'do'

Frequency Distribution

- ▶ Although 1PL seems to have a very low relative frequency too, it is approximately 28 times more frequent than 3PL.

	SG	PL
1	59.2%	1.51%
2	3.06%	3.45%
3	32.7%	0.05%

Table: Averaged frequencies of person/number suffixes in desideratives given one of the verbs *al-* 'buy', *git-* 'go', *gör-* 'see', *yap-* 'do'.

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

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

Implications of Corpus Data

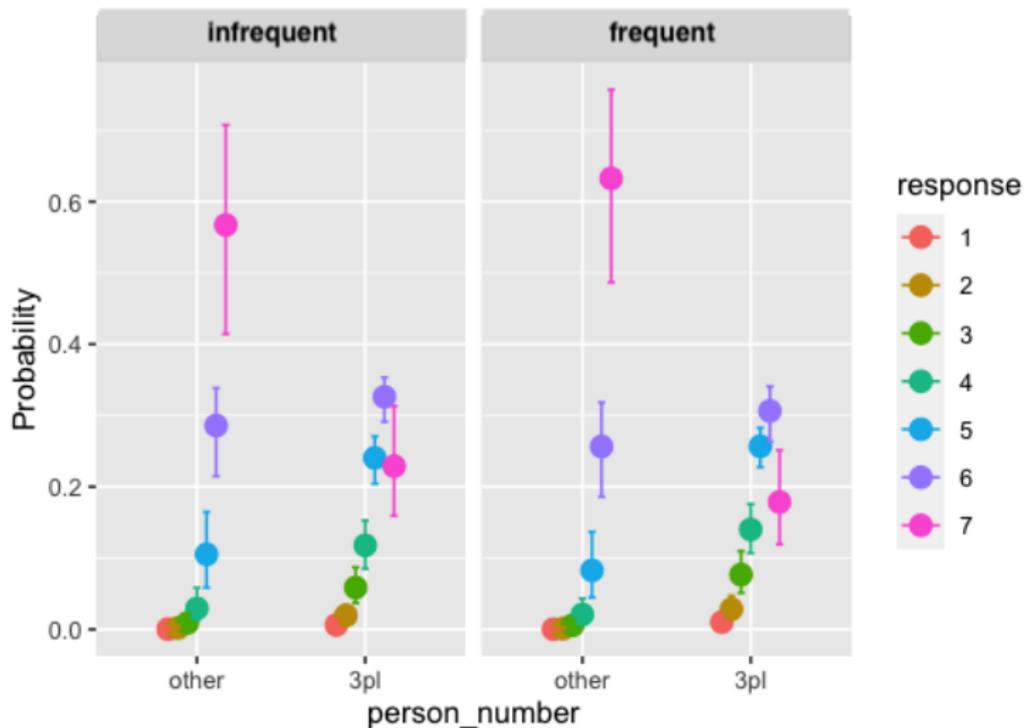
- ▶ 3PL desiderative forms are rarely attested.
- ▶ This cannot be explained by low base frequency of desideratives.
- ▶ There is no semantic reason for why 3PL desideratives are rare (remember *istiyor-*).
- ▶ However, we need a systematic way to extend our search to all desiderative bearing verbs to make sure the frequency distribution is representative of the population of desideratives.

The corpus data is compatible with our intuitions that 3PL desideratives pose a problem for speakers of Turkish. It is abnormally infrequent compared to other forms, suggesting that it instantiates a paradigm gap.

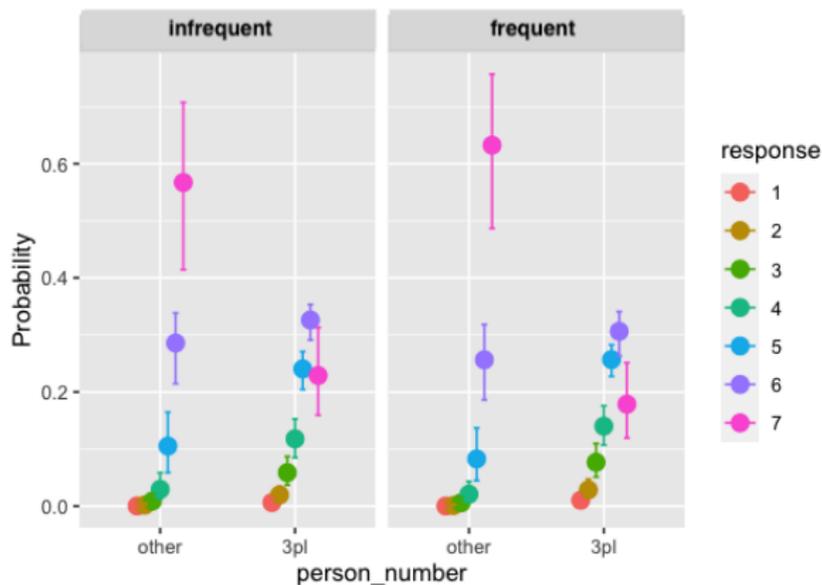
Results

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

$$\text{RESPONSE} \sim 1 + \text{AGR} * \text{FREQ} + (1 + \text{AGR} * \text{FREQ} \mid \text{SUBJECT}) + (1 + \text{AGR} \mid \text{ITEM})$$



Results



- ▶ 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].

Recap

- ▶ We hypothesized that there is a gap in the 3PL cell of the agreement paradigm of desideratives.
- ▶ We illustrated an irregularity in 3SG desideratives and proposed that it has an impact on the (in)ability to derive a 3PL form.
- ▶ We have provided evidence from corpus that 3PL desideratives statistically behave like a paradigm gap even though n is low.
- ▶ Then we discussed experimental data which indicate that 3PL desideratives significantly reduce acceptability compared to desideratives bearing other agreement markers.
- ▶ Also, the model fit to data presents some evidence in favor of the hypothesis that "3PL desideratives reduce acceptability more when they have a frequent stem."

Discussion

- ▶ A competition-based account (Albright & Hayes 2002, Gorman & Yang 2016) can potentially explain the emergence of the gap by assuming that there is no winner at the end of the process where the possible rules for deriving 3PL desideratives compete.

However, it would be hard-pressed to explain why there is a gap instead of free variation.

- ▶ A model of language learning that is sensitive to frequency distributions like that of Daland et.al.'s (2007) can explain how the gap persists given the abnormal frequency distributions of desiderative forms.

But, it would have no means to explain how and why the gap emerged in the first place.

Discussion

- ▶ It is more plausible that both sensitivity to frequency distribution and rule-competition play a role in the existence of the gap, given that one model cannot explain all.
- ▶ Due to the low frequency of -Asl desideratives, it is very likely that there is not enough data to deduce reliable rules from.
- ▶ Without reliable rules, one would feel hesitant to produce a novel form even if they can actually produce it (Albright & Hayes 2003).

Conclusion

- ▶ It is most likely that if one ever tries to produce a 3PL desiderative, it is a real-life wug test (Berko 1958).
- ▶ An experimental wug test is unavoidable. However, we can avoid a real-life wug test and we would avoid it if there is a high risk of failure.
- ▶ Optionality of 3PL agreement suffix in non-pro drop contexts and low base frequency of -Asl desideratives help us avoid it.

Either we fail or we avoid failure.

References

- ▶ Aksan, Y., Aksan M., Mersinli Ü., & Ufuk, U. (2017). A frequency dictionary of Turkish. London: Routledge
- ▶ 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.
- ▶ Bader, M., & Häussler, J. (2009). *Toward a model of grammaticality judgments*. Journal of Linguistics, 45, 1–58.
- ▶ Berko-Gleason, J. (1958). *The child's learning of English morphology*. Word 14. 150–177.
<https://doi.org/10.1080/00437956.1958.11659661>.

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>.

References

- ▶ 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
- ▶ Gorman, K., & Yang, C. (2019). *When nobody wins*. In F. Rainer, F. Gardani, H. C. Luschützky, W. U. Dressler (Eds.), *Competition in inflection and word formation*. Dordrecht: Springer.
- ▶ Göksel, A. (2009). *Compounds in Turkish*. *Lingue e Linguaggio*, 2, 213-236.

References

- ▶ Göksel, A. & Kerslake, C. (2005). *Turkish: A Comprehensive Grammar*. NY: Routledge.
- ▶ 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.
- ▶ Sims, A. (2015). *Inflectional Defectiveness* (Cambridge Studies in Linguistics). Cambridge: Cambridge University Press. doi:10.1017/CBO9781107053854
- ▶ Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX). <https://doi.org/10.17605/OSF.IO/MD832>