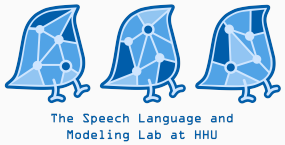


A computational approach to understanding the cognitive reality of morphomic patterns: The case of L-shaped morphomes in Spanish

Akhilesh Kakolu Ramarao[1] Kevin Tang[1,2] Dinah Baer-Henney[3]
{kakolura,kevin.tang,dinah.baer-henney}@hhu.de

August 28th, 2024
LAGB 2024

[1]Heinrich Heine University Düsseldorf, English Language and Linguistics [2]University of Florida, Department of Linguistics [3]Heinrich Heine University Düsseldorf, Institute of Linguistics

Overview

1. Introduction
   - 📚🔍Spanish L-shaped pattern
   - 🧍 The rise and fall of the L-shaped morphome: diachronic and experimental studies (Nevins et al., 2015)
   - 🧍 The cognitive reality of morphomes. Evidence from Italian (Cappellaro et al., 2024)
2. 🖥️Our current study

Introduction

Spanish L-shape pattern

decir	***IND***		***SBJV***	
'to say'	Orthographic	IPA	Orthographic	IPA
1SG	digo	d'igo	diga	d'iga
2SG	dices	d'ises	digas	d'igas
3SG	dice	d'ise	diga	d'iga
1PL	decimos	des'imos	digamos	dig'amos
2PL	decís	des'is	digáis	dig'ajs
3PL	dicen	d'isen	digan	d'igan

Morphomic pattern

- As defined in Mark Aronoff's work "Morphology By Itself" (1994): A morphomic pattern refers to morphological patterns or functions that lack clear motivation from outside the domain of morphology itself.
- A morphological pattern is considered morphomic when its existence cannot be explained by phonological factors (like stem shape) or by the semantic composition of the features it expresses.
- Maiden (2011, 2018, 2021) have identified this morphomic pattern across Romance languages.

The rise and fall of the L-shaped morphome: diachronic and experimental studies (Nevins et al., 2015)

- Investigate the extent to which the “L-shaped morphome" pattern in Romance languages, specifically in Portuguese, Italian, and Spanish, is productive and generalized by native speakers.
- We specifically focus on whether Spanish speakers extend unnatural stem alternation patterns of Spanish verbs to nonce-verbs.

Tú llutes solamente con la mano derecha, pero yo lluso con cualquier mano. Es necesario que ____ con las dos manos para que así seas más productivo.

'You *llutes*2SG.IND only with your right hand, and but I *lluso*1SG.IND with either hand. It's necessary that you ____2SG.SBJ with both hands in order to be more productive.'

- There was a strong preference for Non-L-shaped responses among participants.
- 71.9% of participants (77 out of 107) showed a preference for Non-L-shaped responses over L-shaped responses.
- The study suggests that the “L-shape morphomes" may not be generalized by learners.

The cognitive reality of morphomes. Evidence from Italian

(Cappellaro et al., 2024)

- To replicate and extend the experimental design of Nevins et al. (2015).
- To investigate the cognitive reality and psychological representation of morphomes in Italian verb morphology.

carrier sentence	**Io** *verb*.PRS.IND.1SG **sempre e voi non** *verb*.PRS.IND.2PL **mai**.
	'I always *verb* and you never *verb*.'
task sentence	**Non voglio che anche Maria _____ sempre**.
	'I dont want Maria _____ either.'
target items	*ROOT A verb*.PRS.SBJV.3SG *ROOT B verb*.PRS.SBJV.3SG

- Online[1] forced-choice tasks (with and without eye-tracking).
- Participants chose between L-pattern and non-L-pattern forms for pseudo-verbs.

[1]"online" refers to the fact that the experiment is measuring cognitive processes in real-time

- 63% of participants showed a preference for L-shaped items and 37% for the non-L-items.
- Overall results for response type distribution across experiments converge.
- Native speakers were able to extend the L-pattern that exists in their language to novel words, which is contradictory to the results found in Nevins et al. (2015).

- Inconsistent results between experiments may be attributed to:
    - Methodological differences: production vs. forced-choice tasks
    - Testing environments: Online vs. Offline
    - Linguistic variability: different stimuli across studies
    - Participant factors: limited background data
- In the light of this contradiction, we propose a novel computational approach to simulate these experimental conditions.

Our current study

Main aims

1. To model the learning of the irregular pattern in Spanish using transformer neural networks trained on data with varying frequencies of irregular vs. regular verbs.
2. To analyze the models' performance and error patterns to get insights into how input frequencies and irregularity impact learning of morphomic patterns.
3. To contribute towards understanding the cognitive mechanisms underlying human morphological acquisition and processing, by comparing model behavior to human experimental data in Spanish (Nevins et al., 2015) and Italian (Cappellaro et al., 2024).

Current Study

- To best match the earlier tasks, we performed the morphological reinflection task (Cotterell et al., 2016, 2017, 2018; Vylomova et al., 2020; Pimentel et al., 2021) at the type-level, which is to generate an inflected form from two source form, and their corresponding tags and the target tag, e.g.,
  (source form 1:dices,
  source tag 1:<V;IND;PRS;2;SG>,
  source form 2:digas,
  source tag 2:<V;SBJV;PRS;2;SG>,
  target tag:<V;IND;PRS;1;SG>)
  $\mapsto$ digo. We refer to this input-output combination as a 'triplet'.
- Moreover, three source and four source models did not show a consistent performance improvement over two source models (Liu and Hulden, 2020).

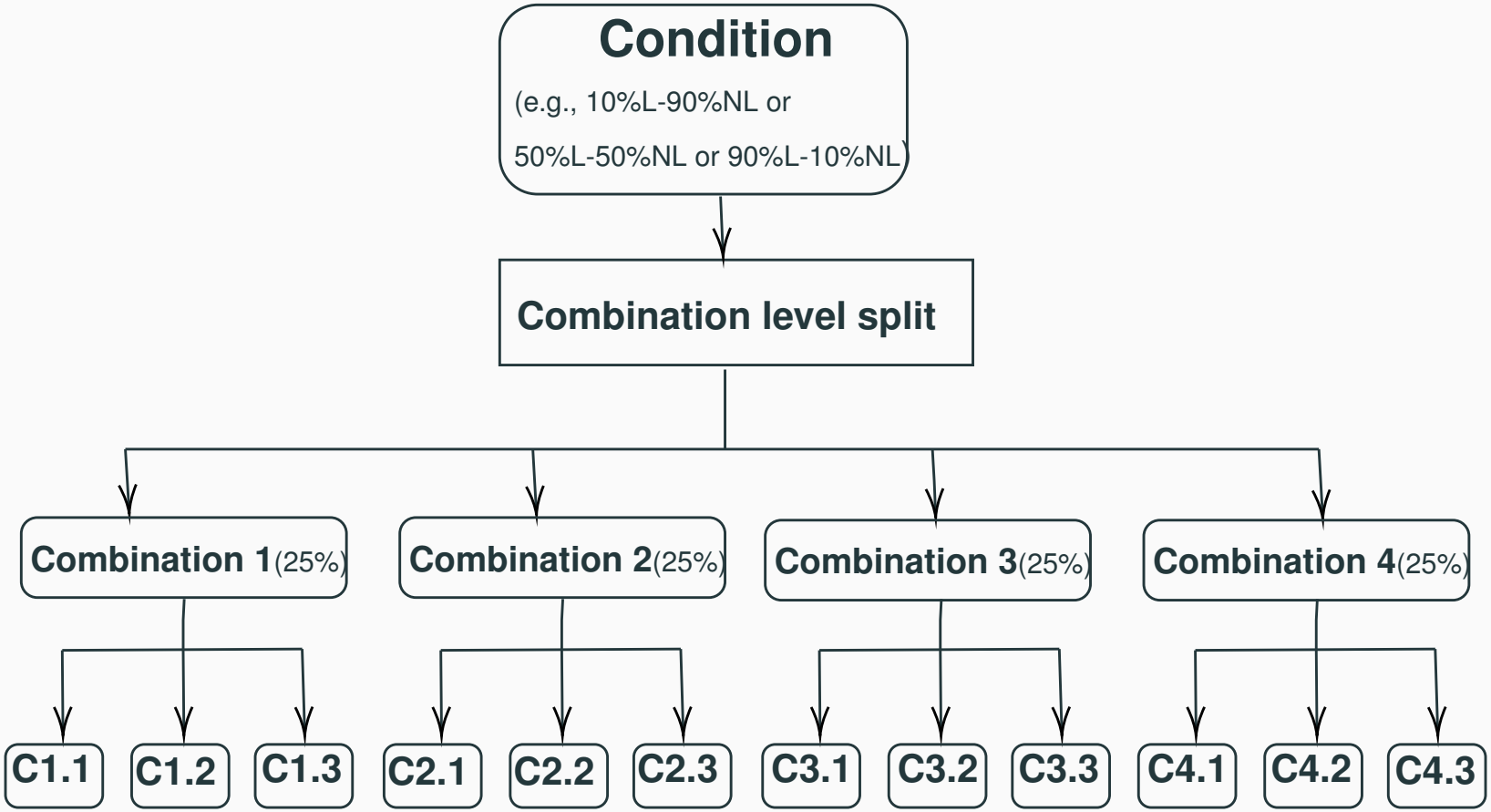
Dataset Construction

- Dataset Source: Spanish corpus from the UniMorph project (Unimorph, 2017)
- Dataset:
  - Initial corpus:
    - 382,956 triplets
    - 5,460 distinct lemmas
  - Filtered Dataset (Indicative and Subjunctive moods only):
    - 165,929 triplets
    - 5,160 distinct lemmas
- L-shaped verbs: 300 lemmas exhibiting the L-shaped pattern
- Non-L-shaped (NL-shaped) verbs: 4,860 lemmas without alternating stems

Experimental Conditions

- **Realistic condition** (10%L-90%NL): Reflects a realistic frequency distribution in the Spanish lexicon.
- Counterfactual conditions designed for direct comparison of learnability between L-shaped and NL-shaped verbs:
  - **Balanced condition** (50%L-50%NL)
  - **Mirror condition** (90%L-10%NL)

Dataset creation workflow



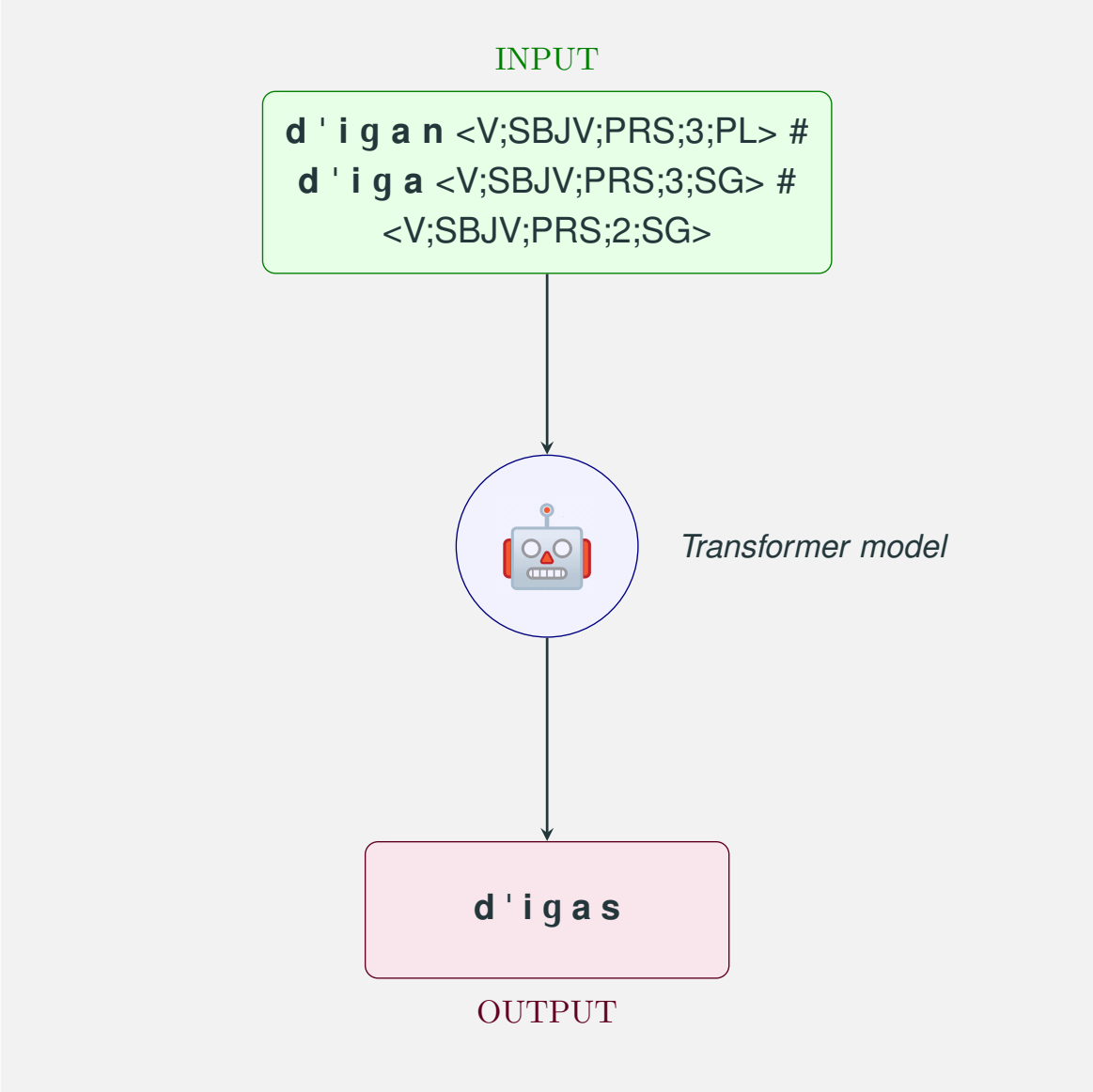
Dataset

- Dataset composition: 38,445 training samples, 5,610 development samples, 43,560 test samples.
- No lemma overlap between training, development, and test sets.
- This hard split strategy is a standard practice in recent studies (Finegan-Dollak et al., 2018; Herzig and Berant, 2020; Lachmy et al., 2022; Goldman et al., 2023) to assess the model's generalization ability.

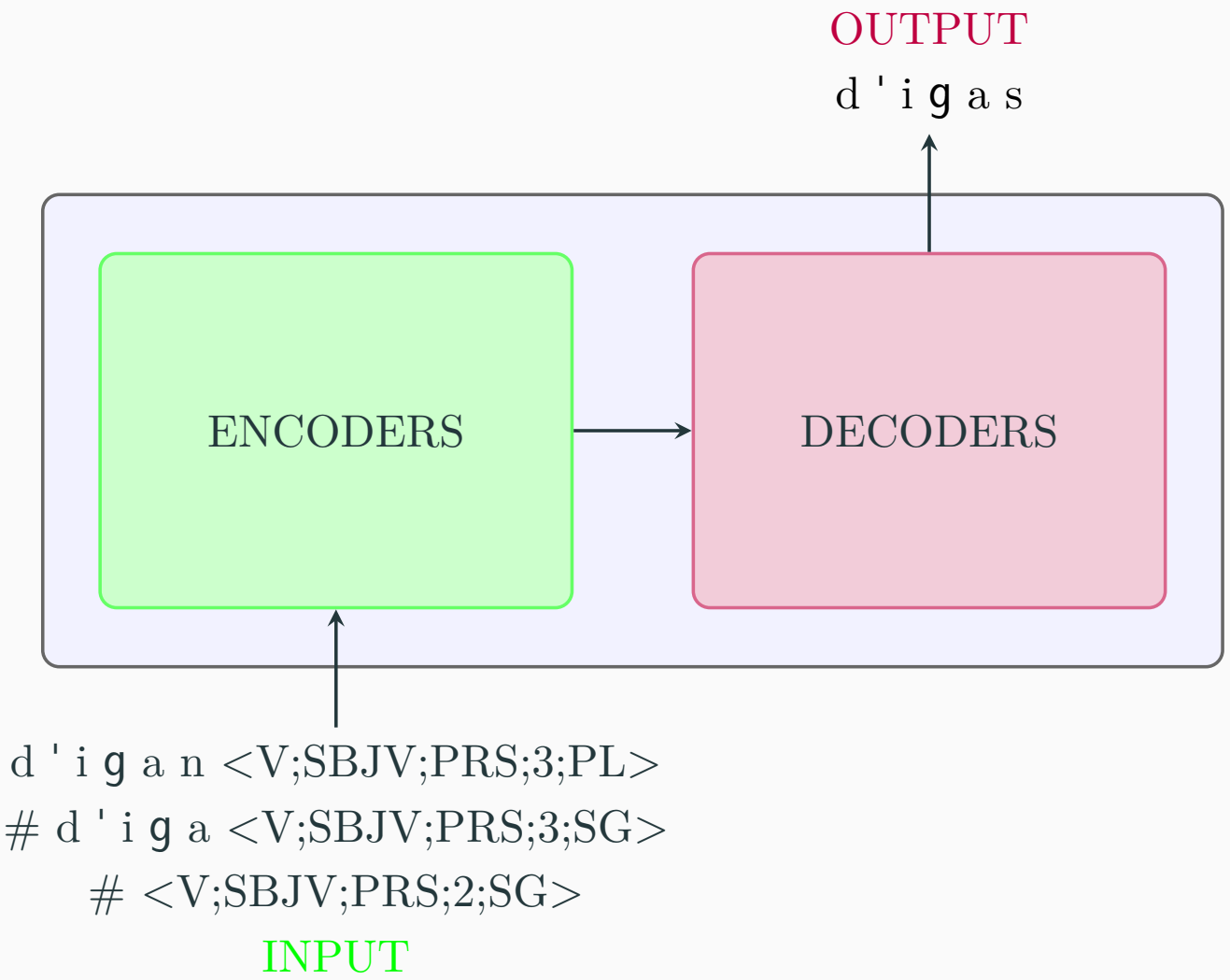
Model Architecture

- Our model is based on the encoder-decoder Transformer for character-level translation, as proposed by Wu et al. (2021).
- This approach has achieved high performance in inflection tasks across many languages (see Cotterell et al. (2017, 2018); Vylomova et al. (2020)).

Model Description



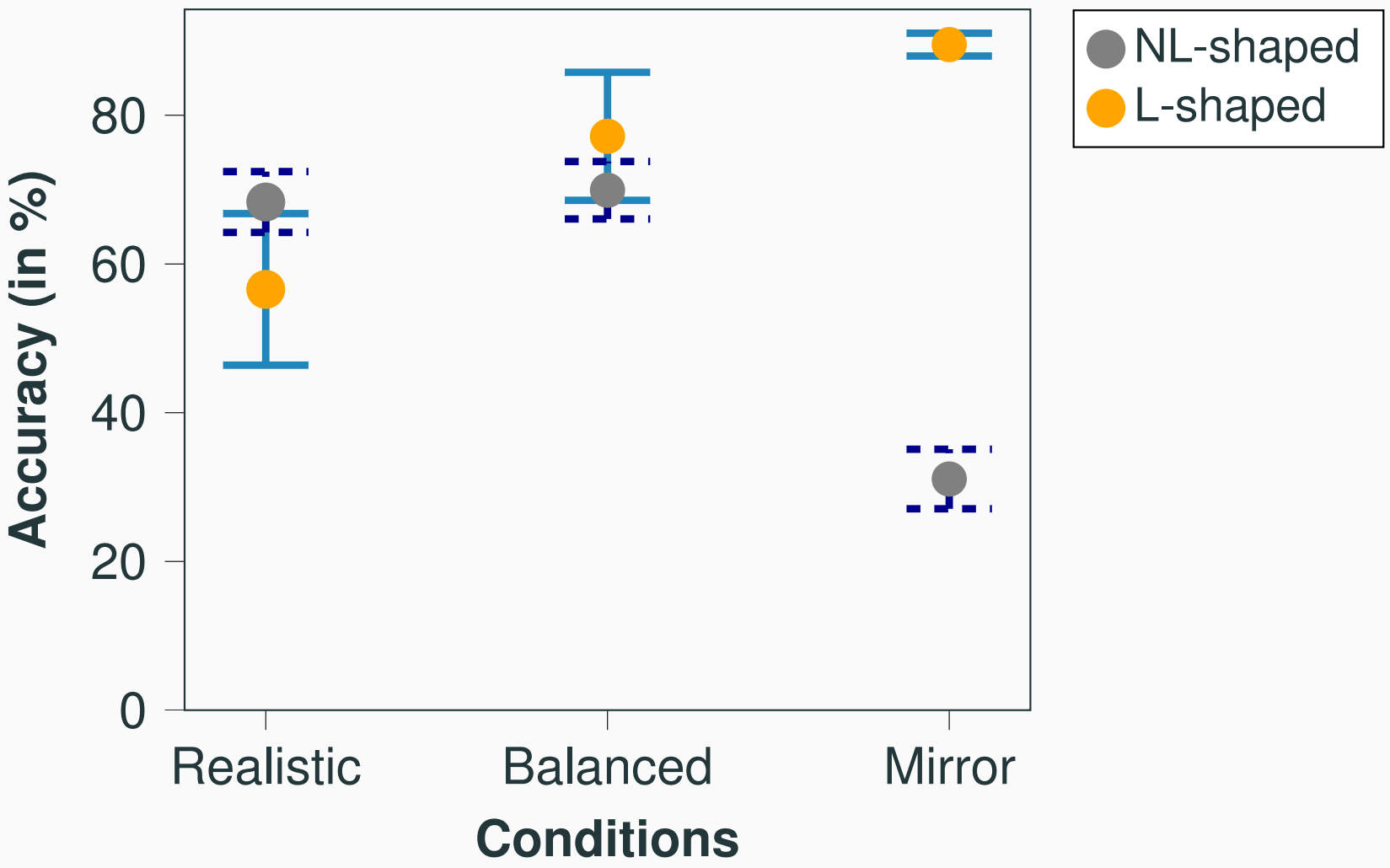
Encoder-decoder Architecture



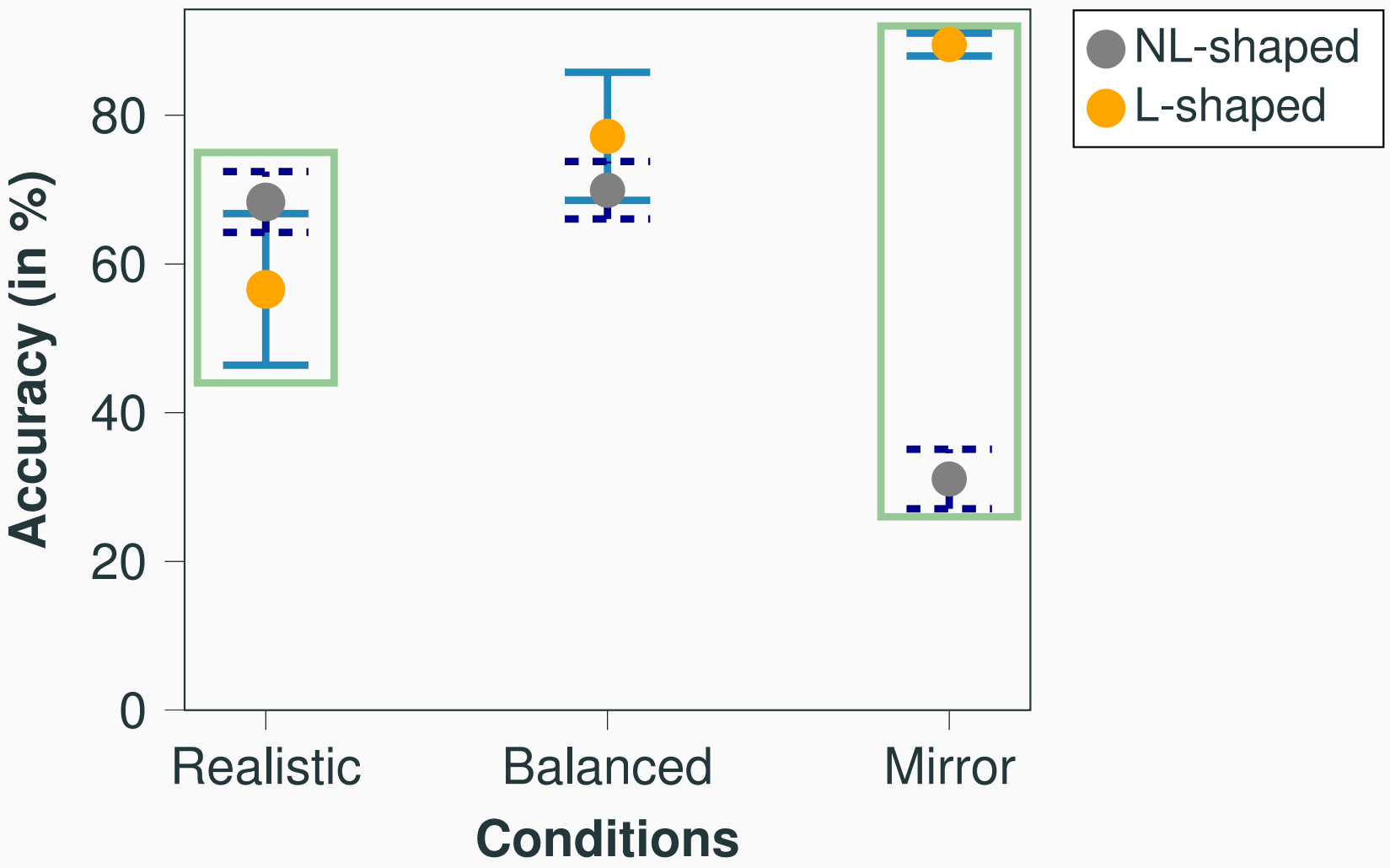
Stem accuracy in realistic condition

Stem accuracy in realistic and balanced conditions

Stem accuracy across conditions



Stem accuracy across conditions



Takeaways

- There is an expected improvement in learning with a higher frequency in the input.
- In the **realistic condition**, the neural network performs better on NL-shaped verbs than on L-shaped verbs. This is consistent with how native Spanish participants performed in the Nevins et al. (2015) study, which showed that the L-shaped pattern is largely unproductive and dispreferred.
- The finding from comparing **realistic** and **mirror** conditions aligns with the results from human forced-choice experiments by Cappellaro et al. (2024), where participants were more likely to extend the L-shaped pattern to novel words even when it was the minority pattern (more than 60% of the times).

Effect of directionality

- The two human experiments checked if direction ($\mathrm{IND}>\mathrm{SBJV}$ or $\mathrm{SBJV}>\mathrm{IND}$) has an effect on the participants choice.
- The effect was not observed in the study by Nevins et al. (2015), but it emerged as a significant main effect in the Cappellaro et al. (2024) study.
- We check how paradigm cell combinations influence the learning of L-shaped verbs.

Results

- The expected type frequency effect can be seen across all cell combinations.
- In the **realistic condition**, NL-shaped verbs were more accurate than L-shaped verbs.
- In the **balanced** and **mirror** conditions, the L-shaped verbs were more accurate than NL-shaped verbs.

Takeaways

- The models' results are in-line with both human experiments.
- We can use use computational methods as a tool to simulate cognition and can serve as a good baseline.
- We can create counter-factual worlds, which can help with hypothesis generation.

Comparison of Studies

	Spanish	**Italian**	**Spanish**
Study	Nevins et al. (2015)	Cappellaro et al. (2024)	Our study
Task type	Production	Forced-choice	Production
Accuracy	33 %	60 %	57 %

References

Chiara Cappellaro, Nina Dumrukcic, Isabella Fritz, Francesca Franzon, and Martin Maiden. 2024. The cognitive reality of morphomes. evidence from italian. *Morphology*, 34(1):33–71.

Ryan Cotterell, Christo Kirov, John Sylak-Glassman, Géraldine Walther, Ekaterina Vylomova, Arya D. McCarthy, Katharina Kann, Sabrina J. Mielke, Garrett Nicolai, Miikka Silfverberg, David Yarowsky, Jason Eisner, and Mans Hulden. 2018. The CoNLL–SIGMORPHON 2018 shared task: Universal morphological reinflection. In *Proceedings of the CoNLL–SIGMORPHON 2018 Shared Task: Universal Morphological Reinflection*, pages 1–27, Brussels. Association for Computational Linguistics.

Ryan Cotterell, Christo Kirov, John Sylak-Glassman, Géraldine Walther, Ekaterina Vylomova, Patrick Xia, Manaal Faruqui, Sandra Kübler, David Yarowsky, Jason Eisner, and Mans Hulden. 2017. CoNLL-SIGMORPHON 2017 shared task: Universal morphological reinflection in 52 languages. In *Proceedings of the CoNLL SIGMORPHON 2017 Shared Task: Universal Morphological Reinflection*, pages 1–30, Vancouver. Association for Computational Linguistics.

Ryan Cotterell, Christo Kirov, John Sylak-Glassman, David Yarowsky, Jason Eisner, and Mans Hulden. 2016. The SIGMORPHON 2016 shared task - morphological reinflection. In *Proceedings of the 14th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology, Berlin, Germany, August 11, 2016*, pages 10–22. Association for Computational Linguistics.

Catherine Finegan-Dollak, Jonathan K Kummerfeld, Li Zhang, Karthik Ramanathan, Sesh Sadasivam, Rui Zhang, and Dragomir Radev. 2018. Improving text-to-sql evaluation methodology. *arXiv preprint arXiv:1806.09029*.

Omer Goldman, Khuyagbaatar Batsuren, Salam Khalifa, Aryaman Arora, Garrett Nicolai, Reut Tsarfaty, and Ekaterina Vylomova. 2023. Sigmorphon–unimorph 2023 shared task 0: Typologically diverse morphological inflection. In *Proceedings of the 20th SIGMORPHON workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 117–125.

Jonathan Herzig and Jonathan Berant. 2020. Span-based semantic parsing for compositional generalization. *arXiv preprint arXiv:2009.06040*.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980.

Royi Lachmy, Valentina Pyatkin, Avshalom Manevich, and Reut Tsarfaty. 2022. Draw me a flower: Processing and grounding abstraction in natural language. *Transactions of the Association for Computational Linguistics*, 10:1341–1356.

Ling Liu and Mans Hulden. 2020. Leveraging principal parts for morphological inflection. In *Proceedings of the 17th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology, SIGMORPHON 2020, Online, July 10, 2020*, pages 153–161. Association for Computational Linguistics.

Martin Maiden. 2011. Allomorphy, autonomous morphology and phonological conditioning in the history of the Daco-Romance present and subjunctive. *Transactions of the Philological Society*, 109(1):59–91.

Martin Maiden. 2018. *The Romance verb: Morphomic structure and diachrony*. Oxford University Press.

Martin Maiden. 2021. The morphome. *Annual Review of Linguistics*, 7:89–108.

Andrew Nevins, Cilene Rodrigues, and Kevin Tang. 2015. The rise and fall of the L-shaped morphome: diachronic and experimental studies. *Probus: International Journal of Latin and Romance Linguistics*, 27(1):101–155.

Tiago Pimentel, Maria Ryskina, Sabrina J. Mielke, Shijie Wu, Eleanor Chodroff, Brian Leonard, Garrett Nicolai, Yustinus Ghanggo Ate, Salam Khalifa, Nizar Habash, Charbel El-Khaissi, Omer Goldman, Michael Gasser, William Lane, Matt Coler, Arturo Oncevay, Jaime Rafael Montoya Samame, Gema Celeste Silva Villegas, Adam Ek, Jean-Philippe Bernardy, Andrey Shcherbakov, Aziyana Bayyr-ool, Karina Sheifer, Sofya Ganieva, Matvey Plugaryov, Elena Klyachko, Ali Salehi, Andrew Krizhanovsky, Natalia Krizhanovsky, Clara Vania, Sardana Ivanova, Aelita Salchak, Christopher Straughn, Zoey Liu, Jonathan North Washington, Duygu Ataman, Witold Kieraś, Marcin Woliński, Totok Suhardijanto, Niklas Stoehr, Zahroh Nuriah, Shyam Ratan, Francis M. Tyers, Edoardo M. Ponti, Grant Aiton, Richard J. Hatcher, Emily Prud'hommeaux, Ritesh Kumar, Mans Hulden, Botond Barta, Dorina Lakatos, Gábor Szolnok, Judit Ács, Mohit Raj, David Yarowsky, Ryan Cotterell, Ben Ambridge, and Ekaterina Vylomova. 2021. SIGMORPHON 2021 shared task on morphological reinflection: Generalization across languages. In *Proceedings of the 18th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 229–259, Online. Association for Computational Linguistics.

Shinsuke Shimojo, Claudiu Simion, Eiko Shimojo, and Christian Scheier. 2003. Gaze bias both reflects and influences preference. *Nature neuroscience*, 6(12):1317–1322.

Unimorph. 2017. Spanish. https://github.com/unimorph/spa/.

Ekaterina Vylomova, Jennifer White, Elizabeth Salesky, Sabrina J. Mielke, Shijie Wu, Edoardo Maria Ponti, Rowan Hall Maudslay, Ran Zmigrod, Josef Valvoda, Svetlana Toldova, Francis Tyers, Elena Klyachko, Ilya Yegorov, Natalia Krizhanovsky, Paula Czarnowska, Irene Nikkarinen, Andrew Krizhanovsky, Tiago Pimentel, Lucas

Torroba Hennigen, Christo Kirov, Garrett Nicolai, Adina Williams, Antonios Anastasopoulos, Hilaria Cruz, Eleanor Chodroff, Ryan Cotterell, Miikka Silfverberg, and Mans Hulden. 2020. SIGMORPHON 2020 shared task 0: Typologically diverse morphological inflection. In *Proceedings of the 17th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 1–39, Online. Association for Computational Linguistics.

Shijie Wu, Ryan Cotterell, and Mans Hulden. 2021. Applying the transformer to character-level transduction. In *EACL*.

Appendix

Appendix: Visualisation of the cell combination accuracies for 10L-90NL condition

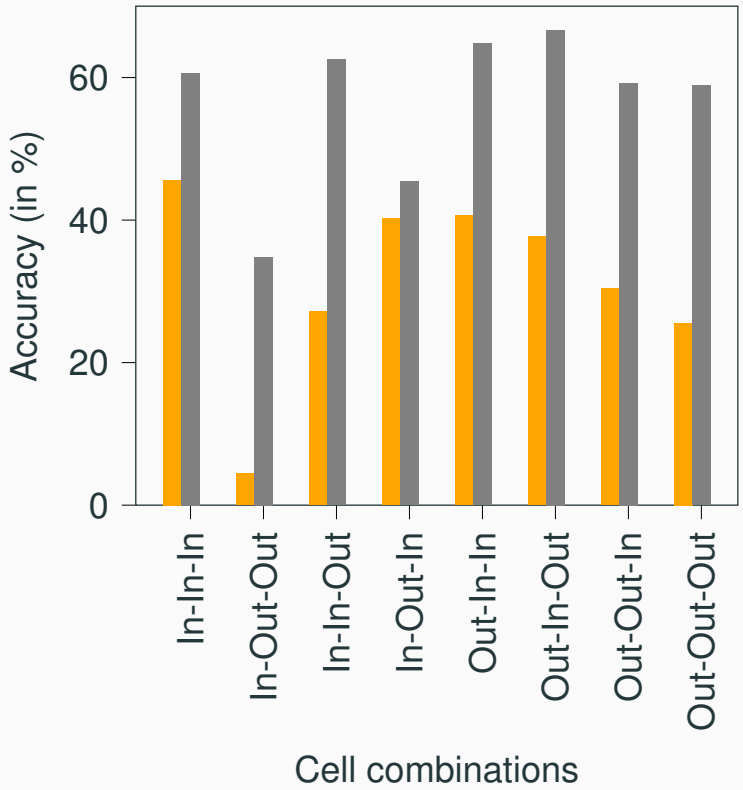


Figure 1: Cell combination accuracies for 10L-90NL condition. Gray: NL-shaped, Orange: L-shaped.

Appendix: Visualisation of the cell combination accuracies for 50L-50NL condition

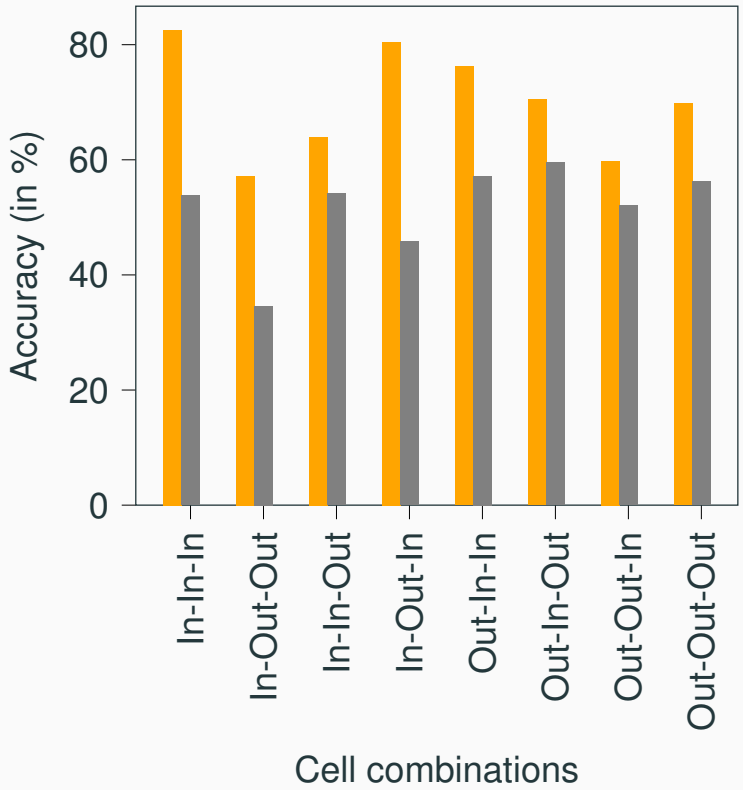


Figure 2: Cell combination accuracies for 50L-50NL condition. Gray: NL-shaped, Orange: L-shaped.

Appendix: Visualisation of the cell combination accuracies for 90L-10NL condition

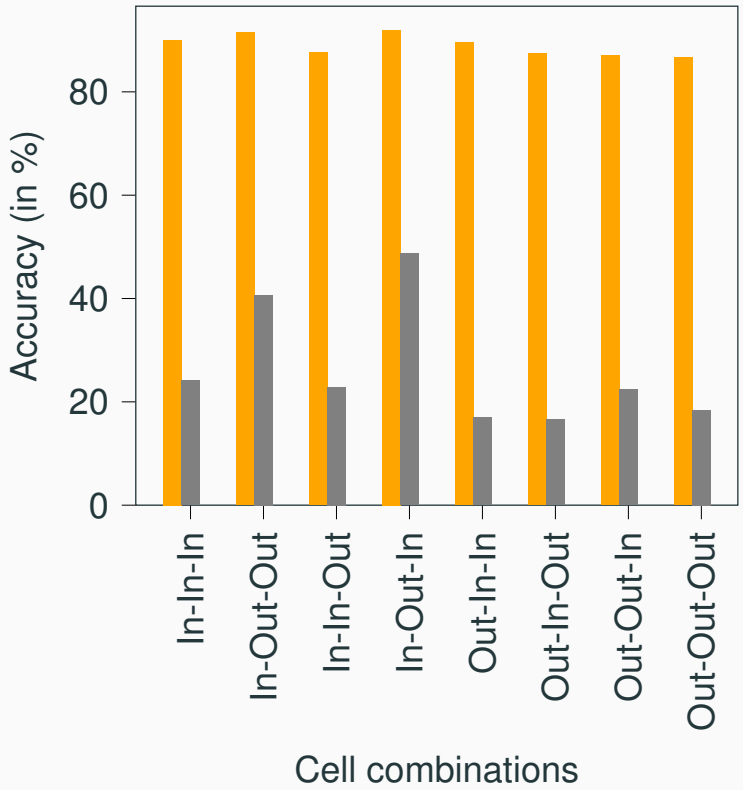


Figure 3: Cell combination accuracies for 90L-10NL condition. Gray: NL-shaped, Orange: L-shaped.

Memorization and Generalization

- **Memorization** is the ratio of number of seen stem-final consonant triples that are correct and the number of attested stem-final consonant triples.
- **Generalization** is the ratio of number of unseen stem-final consonant triples that are correct and the number of unattested stem-final consonant triples.

Memorization and Generalization

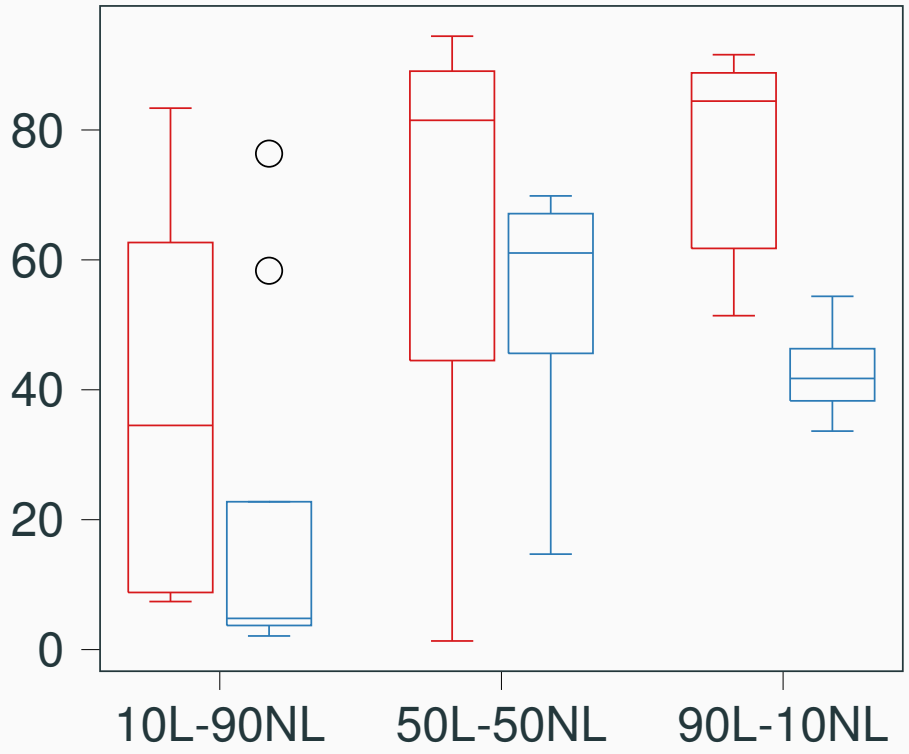


Figure 4: Models' ability to memorize and generalize for stem-final consonant triples that are observed in L-shaped forms across all conditions. Red: Memorization, Blue: Generalization.

Memorization and Generalization

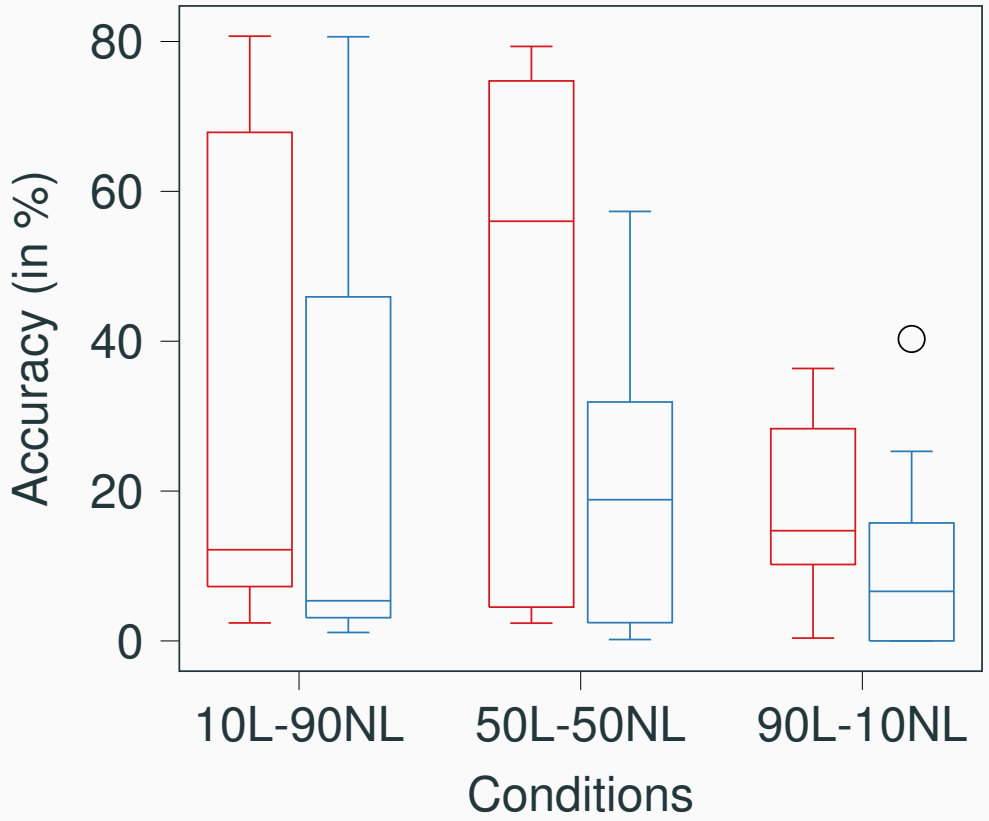


Figure 5: Models' ability to memorize and generalize for stem-final consonant triples that are observed in NL-shaped forms across all conditions. Red: Memorization, Blue: Generalization.

Human Experiment 2

- 135 participants were instructed as follows:
- “You will be presented with examples of invented verbs. Then you will see a sentence in a blank space. Your task is to fill in the blank with the appropriate form of the verb."

- PRS.IND.2PL is presented to the speakers instead of PRS.IND.2SG as in the Nevins et al. (2015).
- For lexical root (IND.PRS.1SG) -
  lg, there is no such verb in the non-first conjugation that does not display a L/U pattern.
- Hypothesis: if the L/U pattern exists in the speakers' mental representation, it must be associated with the inflexion class.
- Root allomorphy exists only in certain phonological alternations: all verbs display a PRS.IND.1SG form with a root ending with a velar g, k,
- They test the hypothesis that the speakers' behaviour may be differentiated by the level of similarity.

- (highest similarity) pseudo-verbs whose root alternation resembled those found in existing L/U- verbs in Italian, labelled as ’mimicking’.
- (low similarity) pseudo-verbs whose root alternation that do not resemble L/U- verbs in Italian, labelled as ’non-mimicking’. This group corresponds to the type of targets used in Nevins et al. (2015).
- (no similarity) pseudo-verbs with extreme root allomorphy.

Experiment 1

carrier sentence	**Io** *verb*.PRS.IND.1SG **sempre e voi non** *verb*.PRS.IND.2PL **mai.** 'I always *verb* and you never *verb.*'
task sentence	**Non voglio che anche Maria ____ sempre.** 'I don't want Maria ____ either'
target items	*ROOT A verb*.PRS.SBJV.3SG *ROOT B verb*.PRS.SBJV.3SG

Figure 6: Stimuli (Image from Cappellaro et al. (2024))

- 31 native Italian speakers
- Each trial involved a first carrier sentence with two forms of the same pseudo-verb followed by a second sentence with a blank space.
- Targets had root allomorphy and participants saw both roots in the carrier sentence (L/U-item with root A and the non-L/U-item with root B).
- The non-L/U-item was in the PRS.IND.2PL.

- They replicated Nevins et al. (2015) in this respect (i.e. Indicative > Subjunctive vs Subjunctive > Indicative directions) with a difference.
- While each participant in the Nevins et al. (2015) study was exposed to only one direction for the whole duration of the experiment, they created a within-subjects design such that half of the participants were shown list 1 where the top 50% of the items had the Indicative > Subjunctive direction and the bottom 50% had Subjunctive > Indicative direction and the other half of the particiapnts saw list 2 with the opoosite direction.

Results

- The highest proportion of L/U-item responses occurred in the mimicking condition (Indicative > Subjunctive direction).
- Next, they checked whether Direction interacted with Shape. There is no overall effect according to whether the target was mimicking, non-mimicking or suppletive.
- They saw a faster reaction times when the L-/U-pattern was chosen.
- When the root-final segment reflects the alternant distribution existing in real verbs the choice of L/U-item tends to to be very high.
- If, however, the segments are distributed between root A and root B in the reverse order, the proportion of L/U-items chosen is considerably lower, in fact even lower than for

- This trend may suggest a weakening effect when an existing phonological alternation is reversed.

Experiment 2: Eye tracking

- ’Gaze-cascade effect’ (Shimojo et al., 2003) - modeling the process of choosing one item over the other when they are simultaneously presented and how this is reflected by looking at that item longer before pressing a key.
- 48 adult Italian readers.

Response type distribution

- The overall response proportion for the L/U-items was 0.63 and 0.37 for the non-L/U-items.
- Overall results for response type distribution across experiment 1 and 2 converge.
- The full mean RTs showed that it took the participants on average longer to press a button when faced with the Indicative > Subjunctive Direction as opposed to the the Subjunctive > Indicative Direction.
- The non-L/U-pattern items had longer reaction times than the L/U-pattern items.

Temporal trajectory of gaze likelihood

- The temporal trajectory of a trial shows the proportion of time a participant is looking at both interest areas (for L/U- and non-L/U-item) at given time intervals during the course of the decision-making process.
- Mean fixation proportion remains relatively consistent for mimicking and non-mimicking items and for the suppletive items there appears to be a large discrepency across bins.

Discussion

- The notable difference between the L/U-item being chosen over the non-L/U-item which is contradictory to the results found in Nevins et al. (2015).
- They indicate that the L/U-morphomic pattern is cognitively real and was activated during the experiment.
- Shape affects the response rate, participants selected the L/U-items more often than the non-L/U-items across all three conditions.
- Direction as a main effect was significant because participants were even more likely manually to select L/U-item when the direction was Indicative > Subjunctive although in both directions the L/U-items were predominantly chosen.

Discussion

- Overall participants took less time to make a decision when they chose L/U-items than when they opted for the non-L/U-item.
- Direction played an important role with the participants taking less time to respond when being prompted with Subjunctive: the L/U-pattern is in fact more likely to be activated quicker in the Subjunctive>Indicative direction compared to the Indicative>Subjunctive direction.
- Within the model for experiment 2, Shape, Choice of L/U-item/non-L/U-item, and Direction as main effects all significantly impacted the reaction times, as did the interactions between Shape:Response Type, however interactions between all three parameters were statistically nonsignificant

Discussion

- The experimental studies were designed to trace the presence of a mental representation of morphomic patterns in Italian native speakers.
- As a starting point, they assessed the robustness of the experimental findings of Nevins et al. (2015).
- Native speakers were able to extend the L/U-pattern that exists in their language to novel words.
- The fact that the suppletive stimuli were processed similarly to mimicking and non-mimicking items is a strong indication that allomorphy is a sufficient (although not necessary) trigger for pattern extension.
- What insights were provided by eyetracking measures with regard to the cognitive processes associated with the

Model Description

- Both Encoder and Decoder have four layers with four attention heads, an embedding size of 256 and hidden layer size of 1,024. We use Adam Optimizer (Kingma and Ba, 2015), with an initial learning rate of 0.001, a batch size of 400, 0.1 label smoothing and 1.0 gradient clip threshold.
- Models are trained for a maximum of 10,000 optimizer updates. Checkpoints are saved every ten epochs. Beam search is used at the decoding time with a beam width of five.

Cell combinations

	10L-90NL			50L-50NL			90L-10NL		
Cell combi.	L	NL	L/NL	L	NL	L/NL	L	NL	L/NL
In-In-In	45.61	60.64	0.7	82.56	53.75	1.63	90.02	24.14	6.85
In-Out-Out	4.44	34.84	0.08	57.18	34.58	1.78	91.53	40.66	2.86
In-In-Out	27.17	62.57	0.4	63.94	54.19	1.15	87.66	22.85	5.07
In-Out-In	40.26	45.4	0.9	80.47	45.83	1.92	91.98	48.72	6.07
Out-In-In	40.62	64.77	0.57	76.17	57.19	1.41	89.54	16.93	15.23
Out-In-Out	37.68	66.68	0.54	70.49	59.58	1.18	87.43	16.53	6.58
Out-Out-In	30.39	59.23	0.45	59.8	52.11	1.08	87.06	22.42	7.44
Out-Out-Out	25.56	58.87	0.38	69.77	56.19	1.23	86.7	18.4	6.19