Frequency matters: Modeling irregular morphological patterns in Spanish with Transformers

Akhilesh Kakolu Ramarao, Kevin Tang, Dinah Baer-Henney
{akhilesh.kakolu.ramarao, kevin.tang, dinah.baer-henney}@uni-duesseldorf.de

Heinrich Heine Universität Düsseldorf

Speech Lexicon and Modeling Lab

ACL 2025 VIENNA
JULY 27 - AUGUST 1

Introduction

🤖 **Model irregular Spanish L-shaped pattern using transformers**

❓ **Compare learning under varying frequency conditions: Irregular patterns are rare – What does a transformer do if they are more frequent?**

🔍 **Analysis of model behavior in terms of:**
- **Overall learning behaviour**
- **Primacy effects**
- **Memorization & generalization capabilities**
- **Consonant alternation sensitivity**

L-shaped Pattern

📖 **Most Spanish verbs are regular with no alternation across the paradigm**

💡 **Few highly frequent verbs follow the L-shaped pattern, where the overall shape is predictable, but the specific alternations are not**

'to eat'	*Indicative*	*Subjunctive*
1SG	como	coma
2SG	comes	comas
3SG	come	coma

Regular verb: ***comer***

'to say'	*Indicative*	*Subjunctive*
1SG	digo	diga
2SG	dices	digas
3SG	dice	diga

L-shaped verb: ***decir***

Vanilla Transformers can learn irregular verbs better than regulars!

Better with L-shaped verbs than with regular verbs, even with low frequency

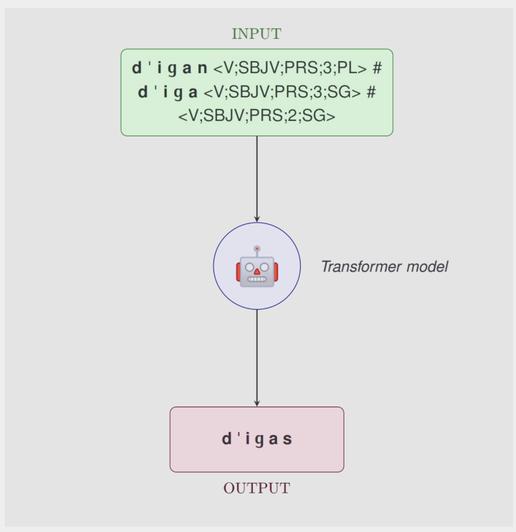
Strong Primacy effect

More L-shaped verbs = More memorization

More frequent the consonant alternation pairs = Better performance

Method

✨ Morphological Re-inflection Task



📚 UniMorph Corpus

📁 300 L-shaped verbs
📋 33 regular verbs
Train: 39,435 | Dev: 4,455 | Test: 44,220

🕹️ Training data comprises full inflection tables, with which the model inflects unseen verbs
✅ No lemma overlap between training and testing data

📊 Experimental Conditions

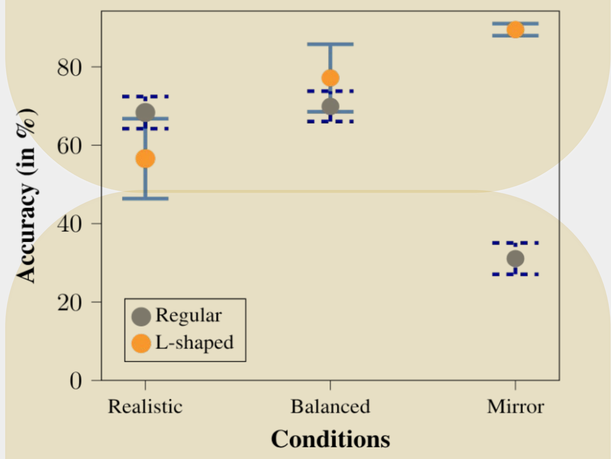
🎯 Realistic : 10%L-90%regular
⚖️ Balanced : 50%L-50%regular
🪞 Mirror : 90%L-10%regular

🚀 Model Architecture

🤖 Encoder-decoder vanilla transformer
🛠️ 12 models per condition
⚙️ Batch size 32 to 3,600

Analysis

Overall learning behaviour

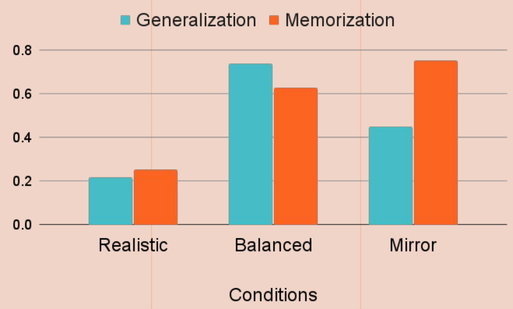


Primacy effects

Cell combi.	Realistic	Balanced	Mirror
In-In-**In** vs. In-In-Out	18.44	18.62	2.36
In-Out-**In** vs. In-Out-Out	35.82	23.29	0.45
Out-In-**Out** vs. In-In-Out	10.51	6.55	-0.23
Out-Out-**Out** vs. Out-Out-In	-4.83	9.97	-0.36

Generalization & memorization

glmer(prediction_status ~ knowledge_state * frequency_condition + (1|triples) + (1|model), data=df, family="binomial")



Consonant alternation sensitivity

🎯 Realistic condition

G\P	[s] - [s]	[s] - [sk]	[s] - [g]	[n] - [ng]
[s] - [sk]	303	**857**	0	0
[s] - [g]	235	21	**99**	0
[n] - [ng]	0	0	0	**281**

⚖️ Balanced condition

G\P	[s] - [sk]	[s] - [g]	[n] - [ng]	[s] - [s]
[s] - [sk]	**6170**	32	0	523
[s] - [g]	112	**613**	0	0
[n] - [ng]	0	0	**853**	0