Morphological Inflection Generation with *Hard Monotonic Attention*

find us on GitHub
roeeaharoni / morphological-reinflection

Roee Aharoni and Yoav Goldberg
Bar Ilan University NLP Lab

1. Abstract

- We present a general-purpose sequence to sequence learning model with a hard-attention mechanism, allowing linear decoding time
- Our model is inspired by the monotonic alignments between the characters in a word and its inflection
- We evaluate the model on several morphological inflection generation datasets, achieving state-of-the-art results in various settings

2. The Task

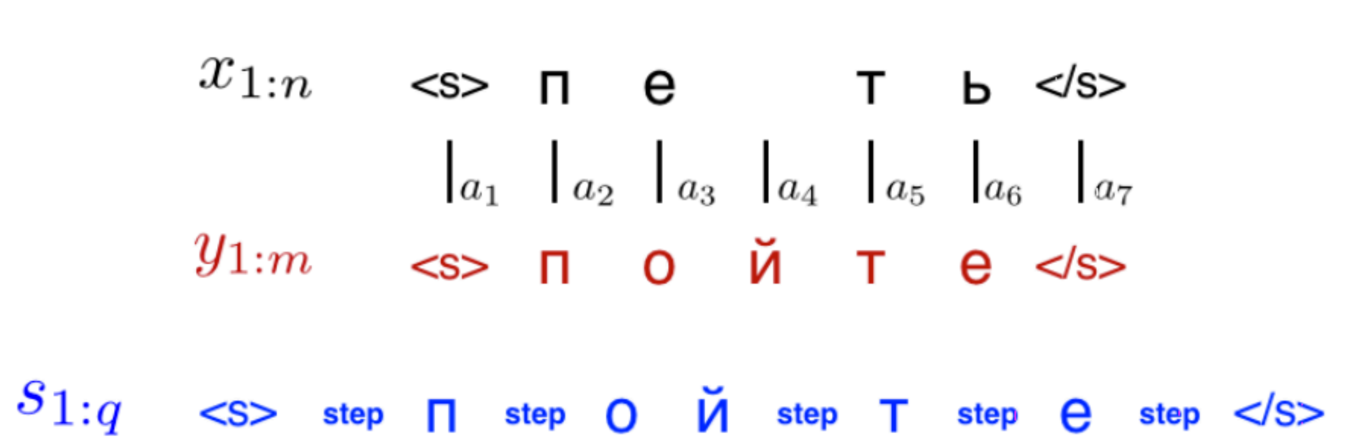
Morphological Inflection Generation involves generating a target word given a **source word** and the **morpho-syntactic attributes** of the target:

input		output
source word	morphological tags	target word
ensamblar	**pos=V, alt=LGSPEC1, mood=SBJV, tense=PST, per=1, num=PL, aspect=IPFV**	**ensambláramos**

3. Previous Approaches

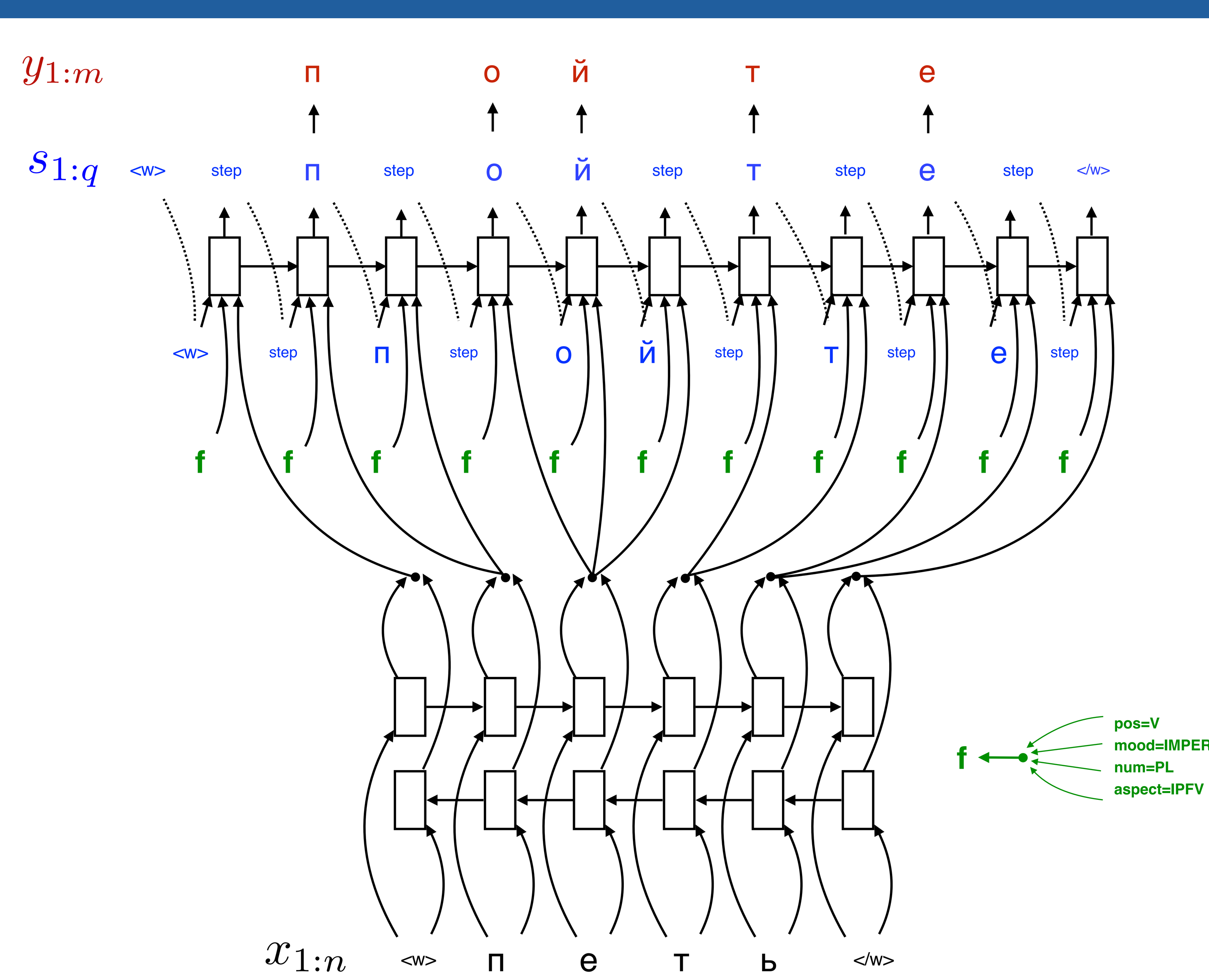
- Vanilla seq2seq (Faruqui et al., 2016) — **not resolution preserving**
- Soft attention (Kann & Schütze, 2016)
  - Decoding time is **quadratic** w.r.t. sequence length
  - Need to jointly learn both alignment and transduction — **hard task given a small amount of examples**

4. From Alignment to Hard Monotonic Attention



- Learn an alignment model over the training data
- Derive an emit/step action sequence from the alignments
- Train a neural network to predict such action sequences given an input sequence

5. The Hard Attention Architecture



6. Results

Small (<500 training examples) - CELEX

	13SIA	2PIE	2PKE	rP	Avg.
MED (Kann and Schütze, 2016a)	83.9	95	87.6	84	87.62
NWFST (Rastogi et al., 2016)	86.8	94.8	87.9	81.1	87.65
LAT (Dreyer et al., 2008)	**87.5**	93.4	87.4	84.9	88.3
Soft	83.1	93.8	88	83.2	87
Hard	85.8	**95.1**	**89.5**	**87.2**	**89.44**

Medium (~13k training examples) - SIGMORPHON 2016

	suffixing+stem changes			circ.	suffixing+agg.+v.h.			c.h.	templatic		
	RU	DE	ES	GE	FI	TU	HU	NA	AR	MA	Avg.
MED	91.46	95.8	98.84	98.5	95.47	98.93	96.8	91.48	**99.3**	**88.99**	95.56
Soft	92.18	96.51	98.88	**98.88**	**96.99**	**99.37**	**97.01**	**95.41**	**99.3**	88.86	**96.34**
Hard	**92.21**	**96.58**	**98.92**	98.12	95.91	97.99	96.25	93.01	98.77	88.32	95.61

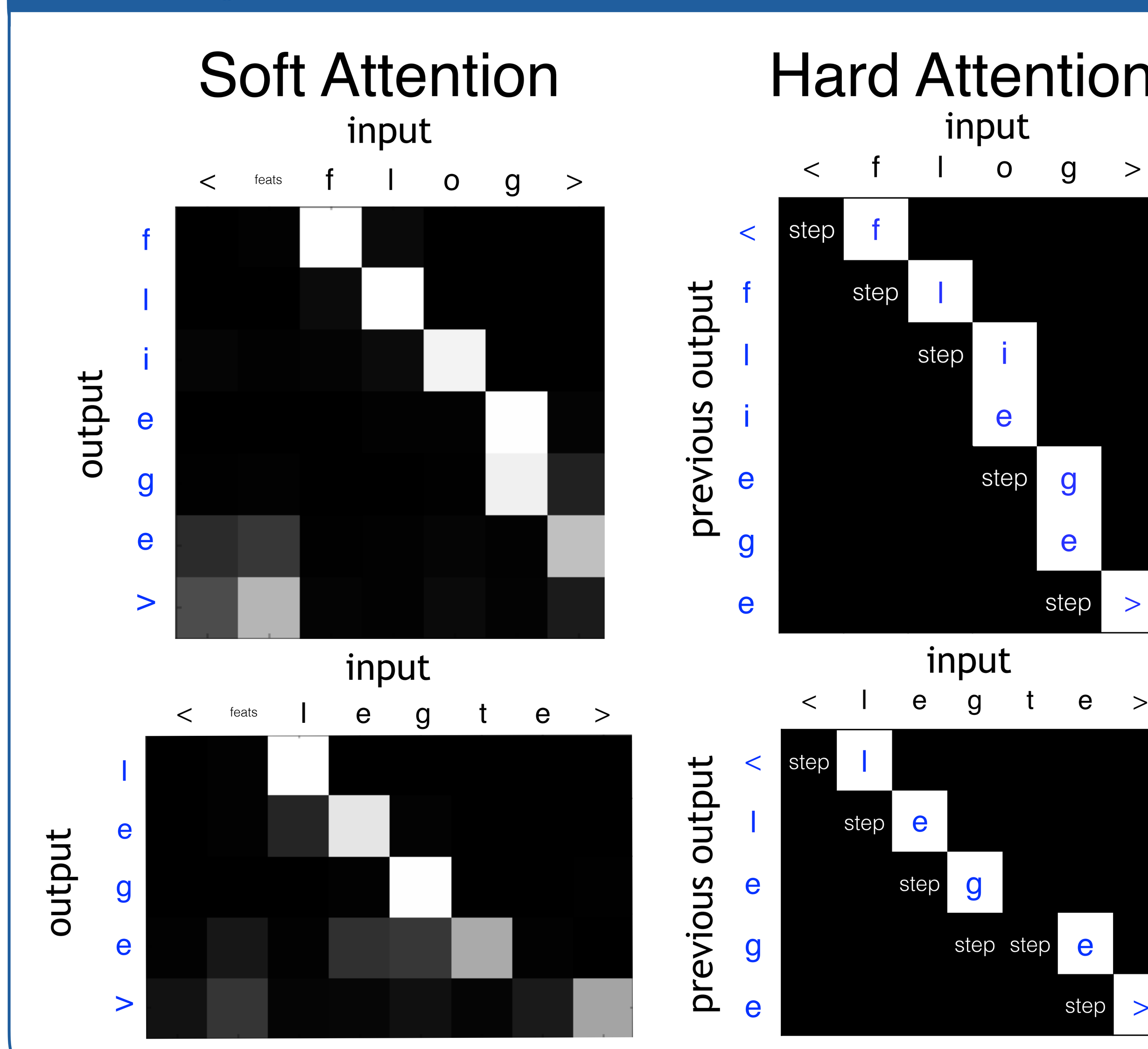
Large (>100k training examples) - Wiktionary

	DE-N	DE-V	ES-V	FI-NA	FI-V	FR-V	NL-V	Avg.
Durrett and DeNero (2013)	88.31	94.76	99.61	92.14	97.23	98.80	90.50	94.47
Nicolai et al. (2015)	88.6	97.50	99.80	93.00	**98.10**	**99.20**	96.10	96.04
Faruqui et al. (2016)	88.12	**97.72**	**99.81**	95.44	97.81	98.82	96.71	96.34
Yu et al. (2016)	87.5	92.11	99.52	95.48	**98.10**	98.65	95.90	95.32
Soft	88.18	95.62	99.73	93.16	97.74	98.79	96.73	95.7
Hard	**88.87**	97.35	99.79	**95.75**	98.07	99.04	**97.03**	**96.55**

7. Recent Adoption

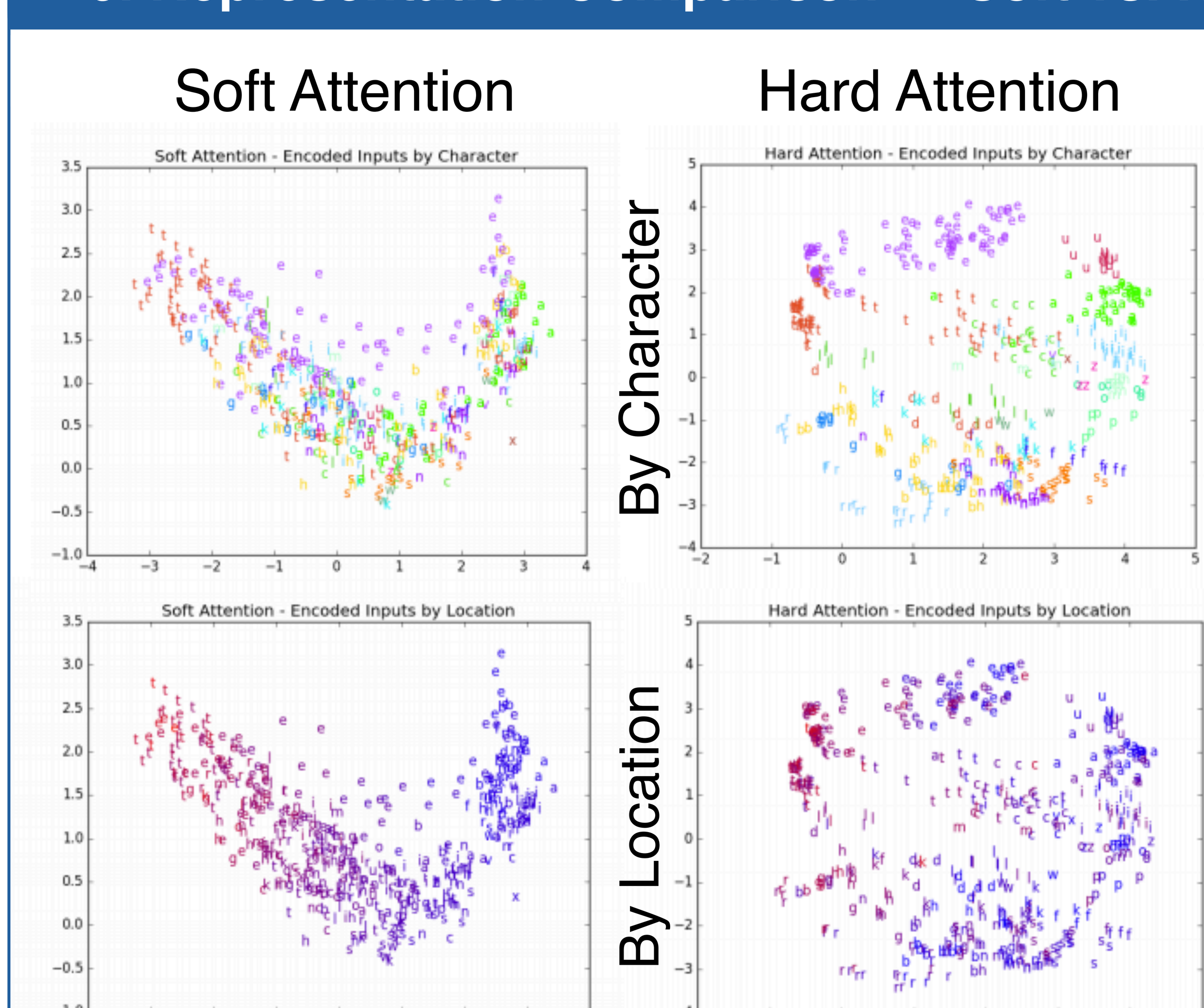
- Makarov et al. (2017) — **Winning system** in CONLL-SIGMORPHON-2017 inflection generation shared task — hard attn. + copying mechanism
- Cotterel et al. (2017) — EACL 2017 outstanding paper — hard attn. + G.M.
- Junczys-Dowmunt & Grundkiewicz (2017) — applied to machine translation automatic post-editing

8. Alignments Comparison — Soft vs. Hard



- Soft attention also learns monotonic alignments
- Hard attention learns one-to-many, deletion using the stepping mechanism
- Hard alignments are more linguistically-sensible

9. Representation Comparison — Soft vs. Hard



- SVD dimension reduction on bidirectional input-character representations
- Soft attention is more sensitive to location
- Hard attention is more sensitive to characters