Better Character Language Modeling Through Morphology

Terra Blevins and Luke Zettlemoyer

UWNLP

Morphologically-Rich Languages Are Hard to Model

	MASKULIN	NEUTRUM	FEMININ	PLURAL
NOMINATIV	neuer ein neuer der neue	neues ein neues das neue	neue eine neue die neue	neue meine neuen die neuen
AKKUSATIV	neuen einen neuen den neuen	neues ein neues das neue	neue eine neue die neue	neue meine neuen die neuen
DATIV	neuem einem neuen dem neuen	neuem einem neuen dem neuen	neuer einer neuen der neuen	neuen +n meinen neuen +n den neuen +n
GENITIV	neuen eines neuen des neuen	neuen eines neuen des neuen	neuer einer neuen der neuen	neuer meiner neuen der neuen

https://www.reddit.com/r/German/comments/71ltao/my_adjective_declension_table/

Morphologically-Rich Languages Are Hard to Model

A word-level LM uses 5 seperate elements of the vocabulary for “neue”

	MASKULIN	NEUTRUM	FEMININ	PLURAL
NOMINATIV	neuer ein neuer der neue	neues ein neues das neue	neue eine neue die neue	neue meine neuen die neuen
AKKUSATIV	neuen einen neuen den neuen	neues ein neues das neue	neue eine neue die neue	neue meine neuen die neuen
DATIV	neuem einem neuen dem neuen	neuem einem neuen dem neuen	neuer einer neuen der neuen	neuen +n meinen neuen +n den neuen +n
GENITIV	neuen eines neuen des neuen	neuen eines neuen des neuen	neuer einer neuen der neuen	neuer meiner neuen der neuen

Morphologically-Rich Languages Are Hard to Model

A word-level LM uses 5 seperate elements of the vocabulary for “neue”

In Finnish, nouns have up to 26 different forms

	MASKULIN	NEUTRUM	FEMININ	PLURAL
NOMINATIV	neuer ein neuer der neue	neues ein neues das neue	neue eine neue die neue	neue meine neuen die neuen
AKKUSATIV	neuen einen neuen den neuen	neues ein neues das neue	neue eine neue die neue	neue meine neuen die neuen
DATIV	neuem einem neuen dem neuen	neuem einem neuen dem neuen	neuer einer neuen der neuen	neuen +n meinen neuen +n den neuen +n
GENITIV	neuen eines neuen des neuen	neuen eines neuen des neuen	neuer einer neuen der neuen	neuer meiner neuen der neuen

Morphologically-Rich Languages Are Hard to Model

A word-level LM uses 5 seperate elements of the vocabulary for “neue”

In Finnish, nouns have up to 26 different forms

Character-level LMs allow information sharing between similar words

	MASKULIN	NEUTRUM	FEMININ	PLURAL
NOMINATIV	neuer ein neuer der neue	neues ein neues das neue	neue eine neue die neue	neue meine neuen die neuen
AKKUSATIV	neuen einen neuen den neuen	neues ein neues das neue	neue eine neue die neue	neue meine neuen die neuen
DATIV	neuem einem neuen dem neuen	neuem einem neuen dem neuen	neuer einer neuen der neuen	neuen +n meinen neuen +n den neuen +n
GENITIV	neuen eines neuen des neuen	neuen eines neuen des neuen	neuer einer neuen der neuen	neuer meiner neuen der neuen

Corpora Have Sparse Coverage of Inflected Forms

Corpora Have Sparse Coverage of Inflected Forms

% of Forms not covered by Train Set

FR: 27% of dev set

RU: 30% of dev set

FI: 46% of dev set

Corpora Have Sparse Coverage of Inflected Forms

% of Forms not covered by Train Set

EN: 27% of dev set

RU: 30% of dev set

FI: 46% of dev set

Prior work shows that highly inflected languages are more difficult to model with a character LM (Cotterell et al., 2018)

Ryan Cotterell et al. Are all languages equally hard to language-model? In *NAACL*, 2018.

Problem: character LMs have capacity to model morphologically regularities, but struggle to capture them from raw text

Problem: character LMs have capacity to model morphologically regularities, but struggle to capture them from raw text

Solution? adding morphology features as objectives to character LM

Approach

$\mathbf{c} = c_1, c_2, ..., c_n$

$\mathbf{m} = m_1, m_2, ..., m_n$

Approach

$$\mathbf{c} = c_1, c_2, ..., c_n$$

$$\mathbf{m} = m_1, m_2, ..., m_n$$

Probability of character c_{t+1}

$$p(c_{t+1}|c_{\leq t}) = \mathrm{softmax}(\mathrm{LSTM}(\mathbf{w}_t, \mathbf{h}_{t-1}))$$

Approach

$$\mathbf{c} = c_1, c_2, ..., c_n \qquad \mathbf{m} = m_1, m_2, ..., m_n$$

Probability of character c_{t+1}

$$p(c_{t+1}|c_{\leq t}) = \mathrm{softmax}(\mathrm{LSTM}(\mathbf{w}_t, \mathbf{h}_{t-1}))$$

Language modeling objective

$$\mathcal{L}_{\mathrm{LM}}(\mathbf{c}) = \mathrm{NLL}(\mathbf{c}) = -\sum_{t=1}^{|\mathbf{c}|} \log p(c_t|c_{<t})$$

Approach

$$\mathbf{c} = c_1, c_2, ..., c_n \qquad \mathbf{m} = m_1, m_2, ..., m_n$$

Probability of character c_{t+1}

$$p(c_{t+1}|c_{\leq t}) = \mathrm{softmax}(\mathrm{LSTM}(\mathbf{w}_t, \mathbf{h}_{t-1}))$$

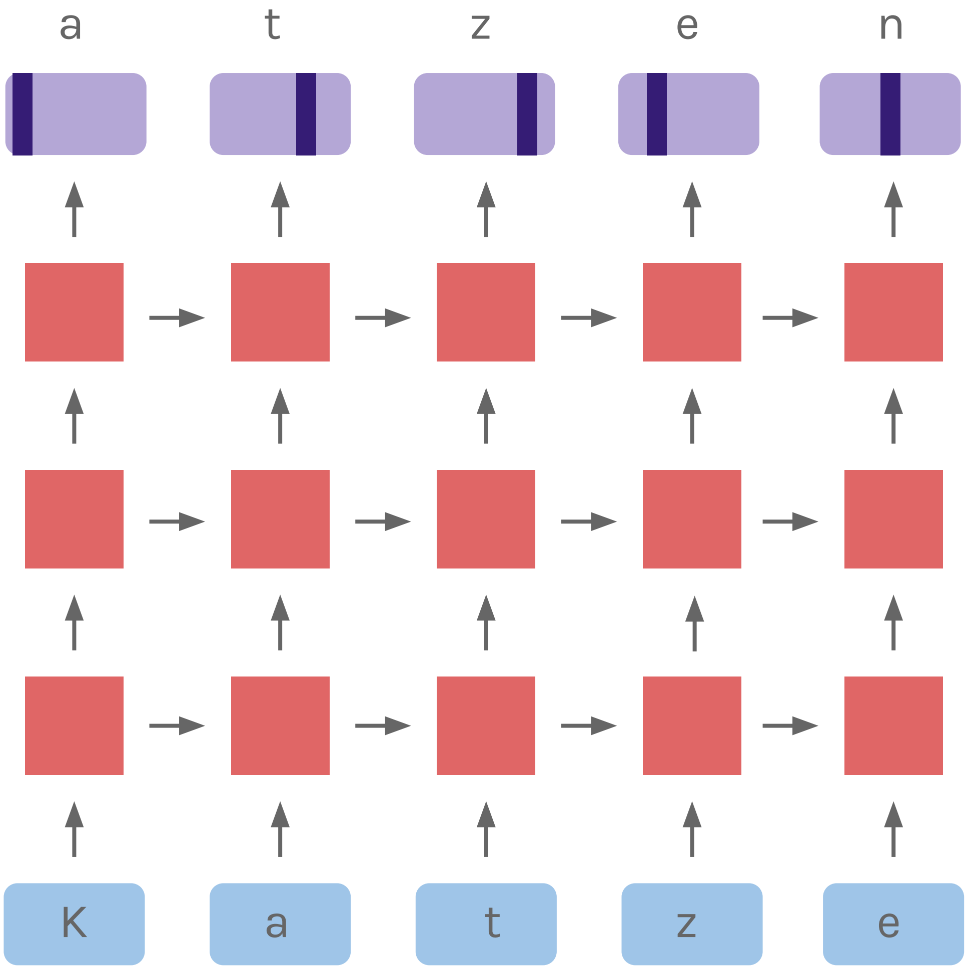
Language modeling objective

$$\mathcal{L}_{\mathrm{LM}}(\mathbf{c}) = \mathrm{NLL}(\mathbf{c}) = -\sum_{t=1}^{|\mathbf{c}|} \log p(c_t|c_{<t})$$

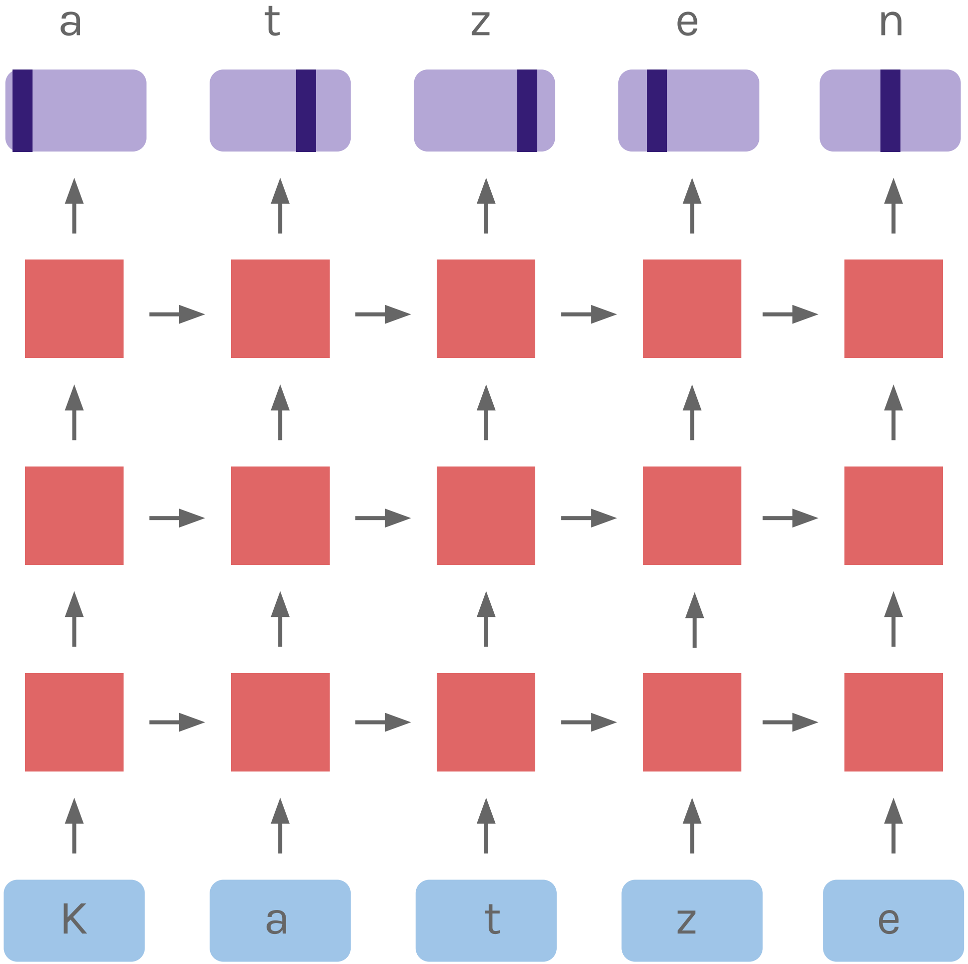
Multitask learning objective

$$\mathcal{L}(\mathbf{c}, \mathbf{m}) = \mathcal{L}_{\mathrm{LM}}(\mathbf{c}) + \delta \sum_{i=1}^{n} \mathcal{L}_i(\mathbf{m})$$

Model Architecture

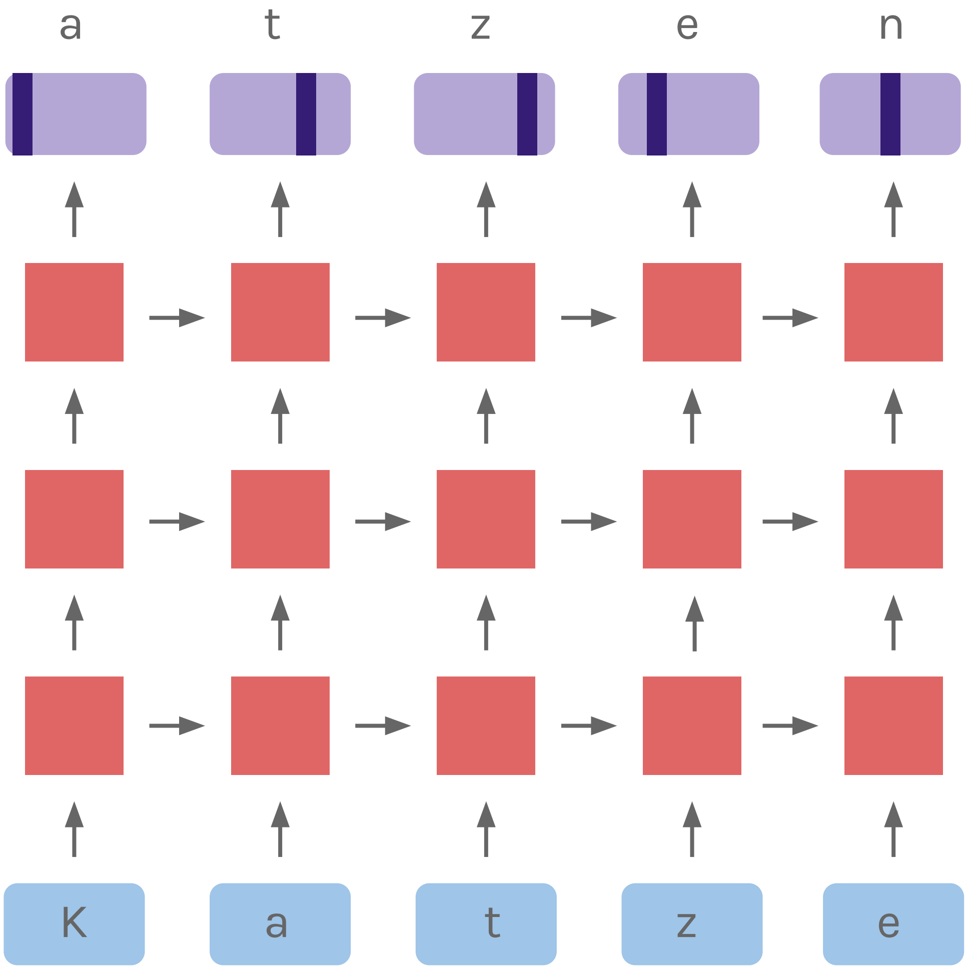


Model Architecture

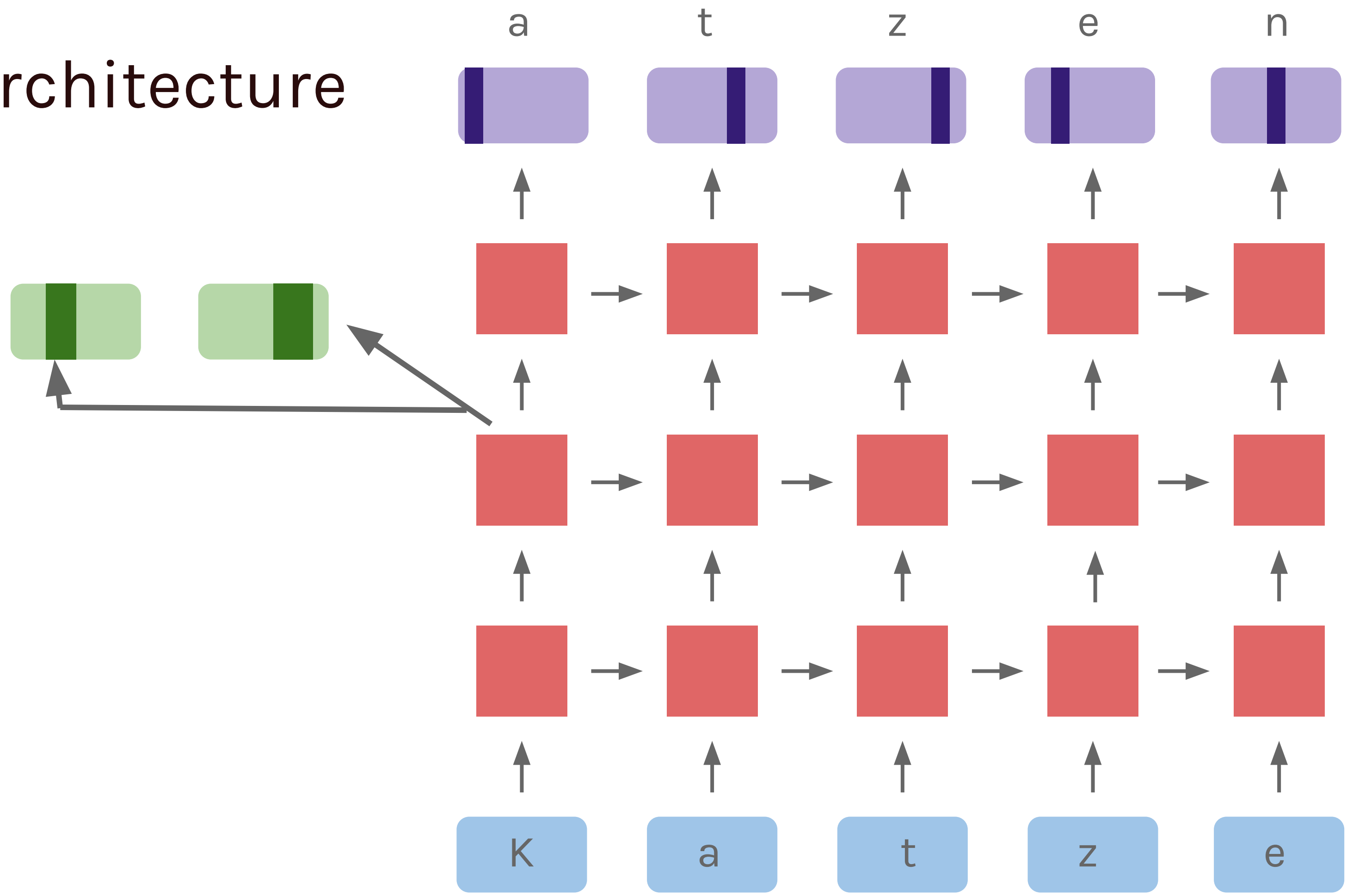


Model Architecture

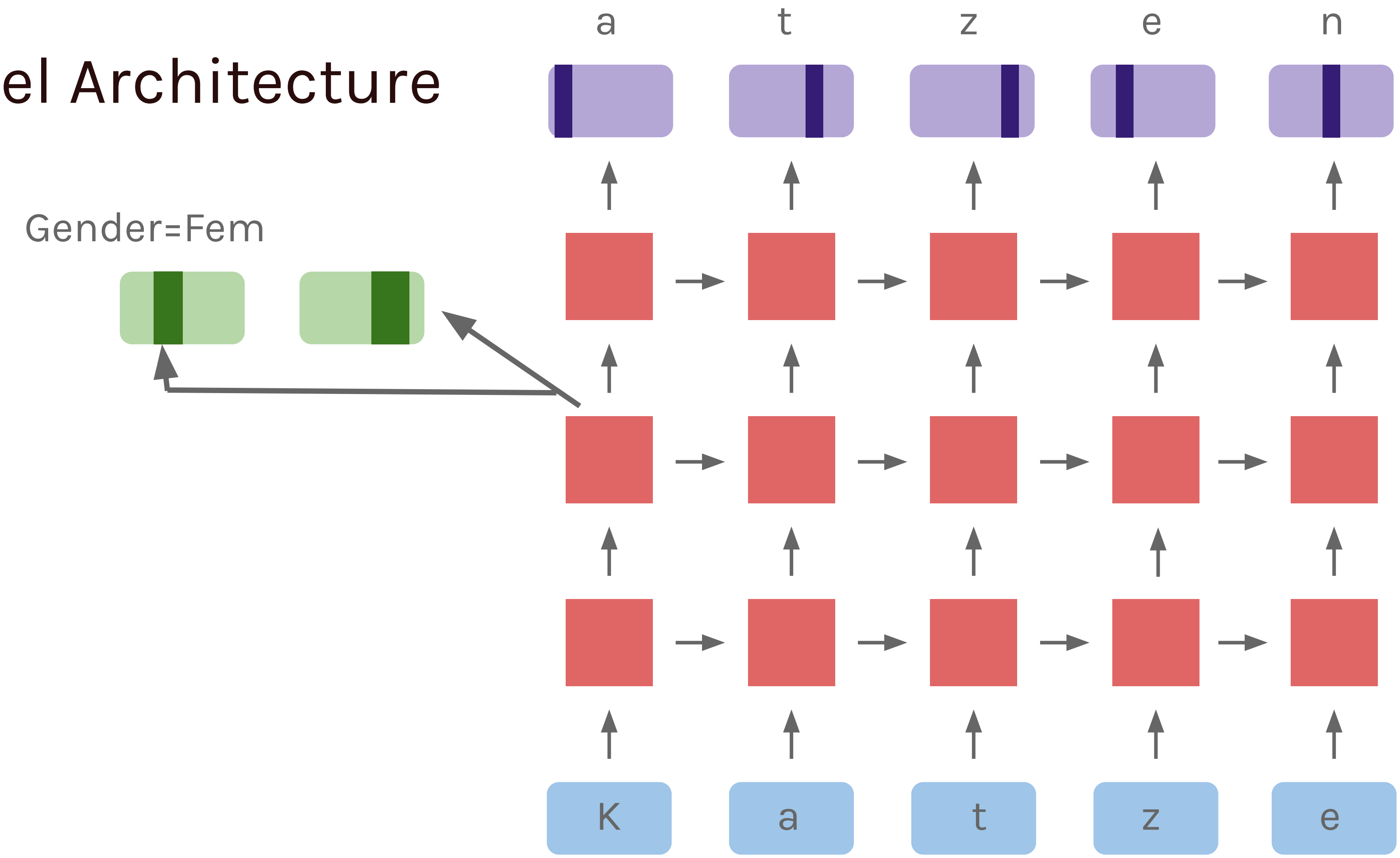
Baseline Character LM



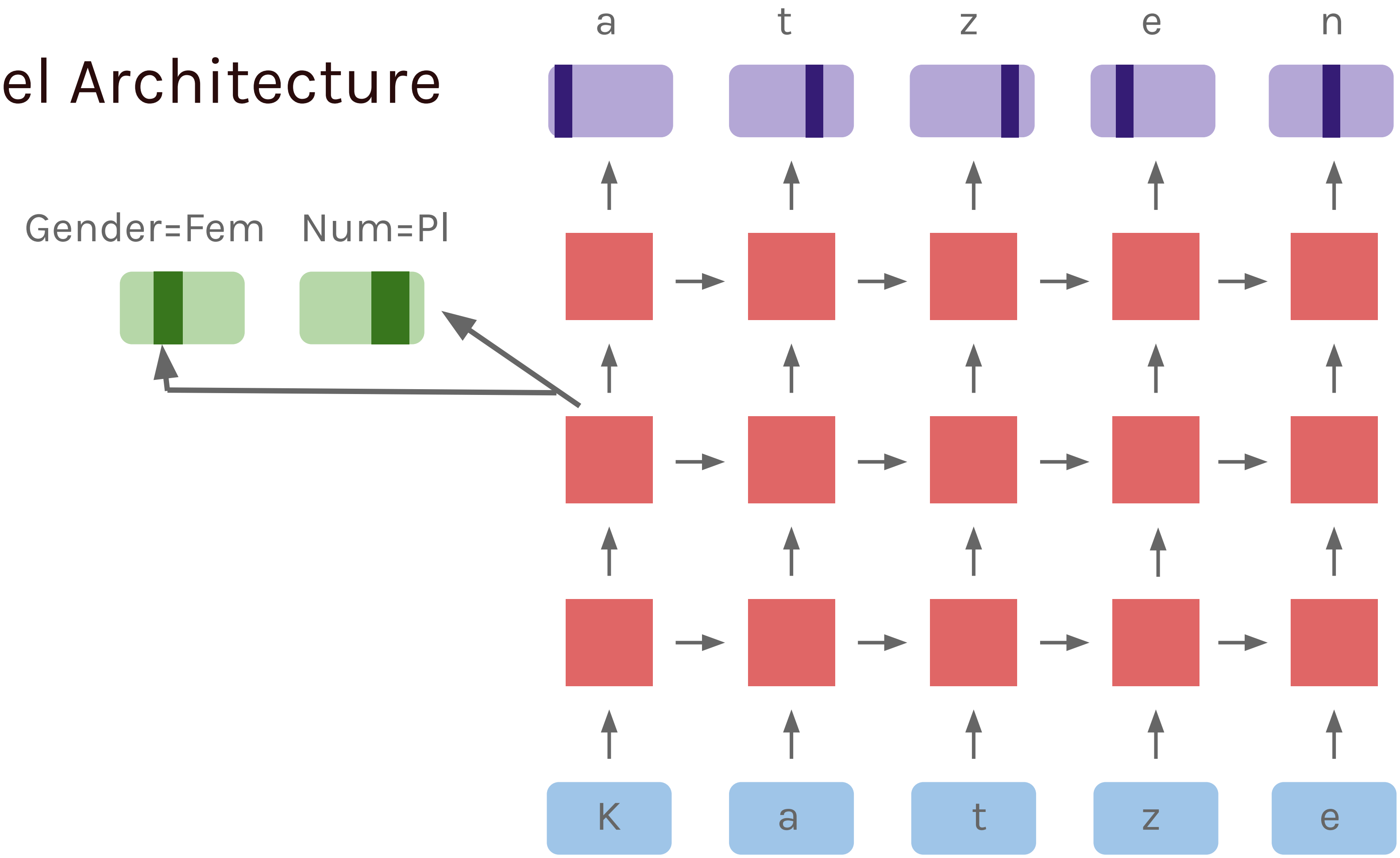
Model Architecture



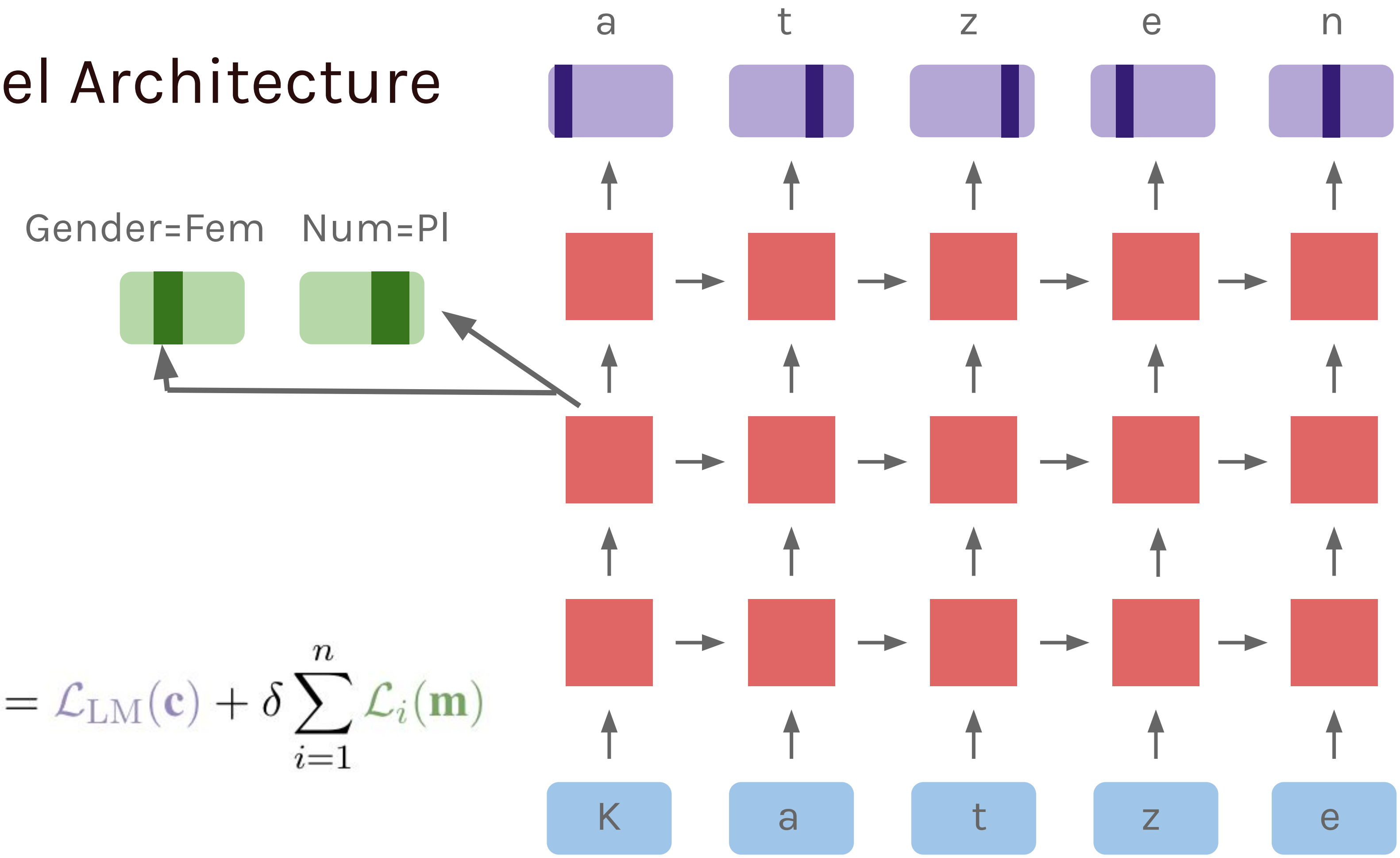
Model Architecture



Model Architecture

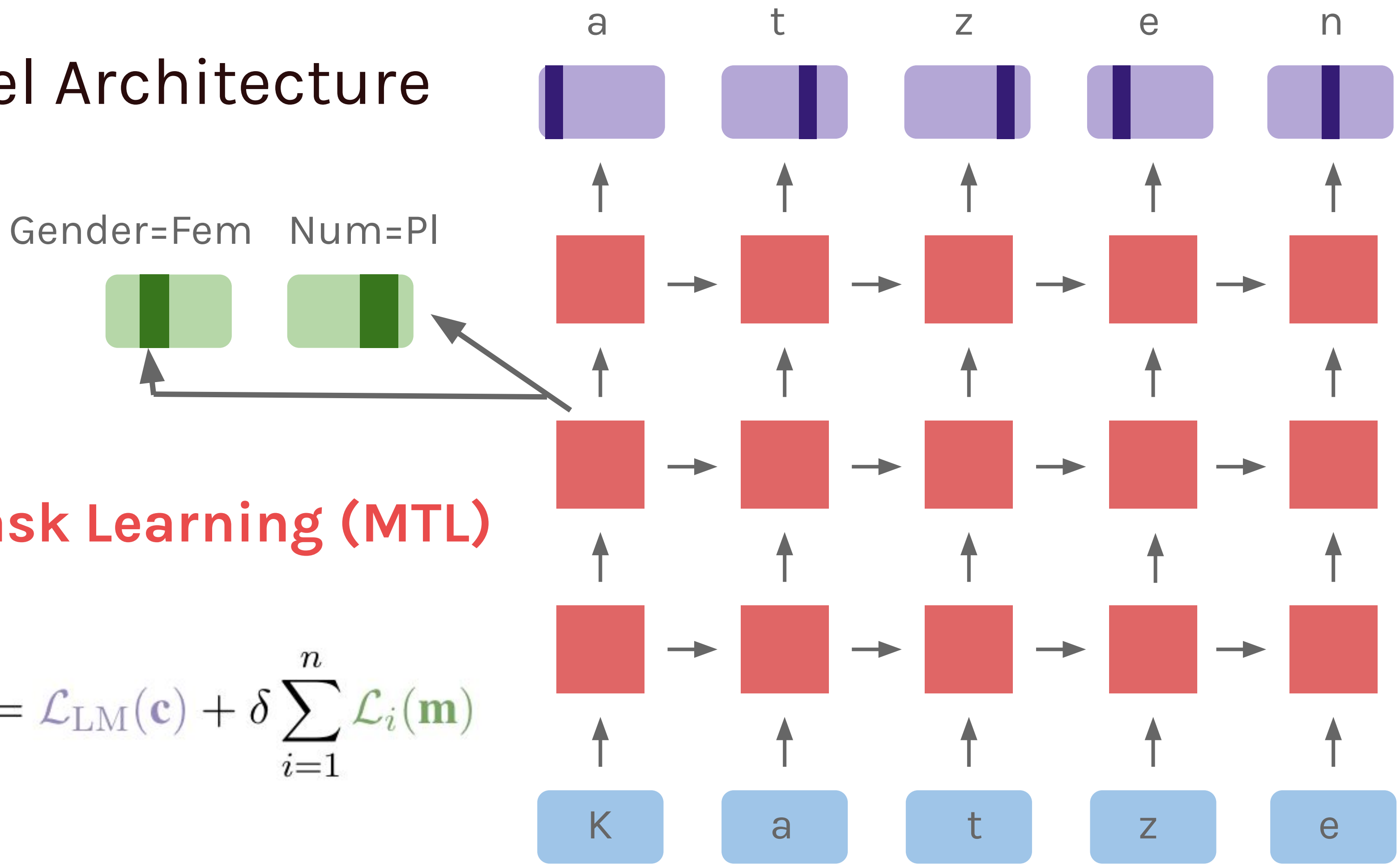


Model Architecture



$$\mathcal{L}(\mathbf{c}, \mathbf{m}) = \mathcal{L}_{\mathrm{LM}}(\mathbf{c}) + \delta \sum_{i=1}^{n} \mathcal{L}_i(\mathbf{m})$$

Model Architecture



Multitask Learning (MTL)

$$\mathcal{L}(\mathbf{c},\mathbf{m}) = \mathcal{L}_{\mathrm{LM}}(\mathbf{c}) + \delta \sum_{i=1}^{n} \mathcal{L}_i(\mathbf{m})$$

Language Modeling: Fully Supervised Setting

CLMs trained with Universal Dependencies for both LM, morphology supervision

Lang	ISO	% Infl	LM	MTL	Δ
Bulgarian	BG	39%	1.890	1.887	0.003
Catalan	CA	31%	1.653	1.599	0.054
Czech	CS	43%	2.045	1.832	0.213
Danish	DA	30%	2.152	2.135	0.017
German	DE	33%	1.917	1.881	0.036
English	EN	15%	2.183	2.173	0.010
Spanish	ES	28%	1.801	1.763	0.038
Farsi	FA	27%	2.213	2.205	0.008
French	FR	32%	1.751	1.712	0.039
Hindi	HI	28%	1.819	1.773	0.046
Croatian	HR	49%	1.866	1.841	0.025
Italian	IT	36%	1.595	1.554	0.041
Latvian	LV	47%	2.243	2.217	0.026
Dutch	NL	19%	1.989	1.972	0.017
Polish	PL	42%	2.218	2.154	0.064
Portuguese	PT	31%	1.787	1.785	0.002
Romanian	RO	42%	1.840	1.798	0.042
Russian	RU	42%	1.993	1.824	0.169
Slovak	SK	45%	2.705	2.686	0.019
Ukranian	UK	40%	2.359	2.338	0.021
Estonian	ET	49%	2.089	1.993	0.096
Finnish	FI	55%	1.981	1.921	0.060
Arabic	AR	86%	1.724	1.708	0.016
Hebrew	HE	42%	2.293	2.282	0.011

Language Modeling: Fully Supervised Setting

CLMs trained with Universal Dependencies for both LM, morphology supervision

$$\mathrm{BPC}(\mathbf{c}) = -\frac{1}{|\mathbf{c}|}\sum_{i=1}^{|\mathbf{c}|}\log p(c_i|c_{<i})$$

Lang	ISO	% Infl	LM	MTL	Δ
Bulgarian	BG	39%	1.890	1.887	0.003
Catalan	CA	31%	1.653	1.599	0.054
Czech	CS	43%	2.045	1.832	0.213
Danish	DA	30%	2.152	2.135	0.017
German	DE	33%	1.917	1.881	0.036
English	EN	15%	2.183	2.173	0.010
Spanish	ES	28%	1.801	1.763	0.038
Farsi	FA	27%	2.213	2.205	0.008
French	FR	32%	1.751	1.712	0.039
Hindi	HI	28%	1.819	1.773	0.046
Croatian	HR	49%	1.866	1.841	0.025
Italian	IT	36%	1.595	1.554	0.041
Latvian	LV	47%	2.243	2.217	0.026
Dutch	NL	19%	1.989	1.972	0.017
Polish	PL	42%	2.218	2.154	0.064
Portuguese	PT	31%	1.787	1.785	0.002
Romanian	RO	42%	1.840	1.798	0.042
Russian	RU	42%	1.993	1.824	0.169
Slovak	SK	45%	2.705	2.686	0.019
Ukranian	UK	40%	2.359	2.338	0.021
Estonian	ET	49%	2.089	1.993	0.096
Finnish	FI	55%	1.981	1.921	0.060
Arabic	AR	86%	1.724	1.708	0.016
Hebrew	HE	42%	2.293	2.282	0.011

Language Modeling: Fully Supervised Setting

CLMs trained with Universal Dependencies for both LM, morphology supervision

MTL improves over LM baseline on all 24 languages

$$\mathrm{BPC}(\mathbf{c}) = -\frac{1}{|\mathbf{c}|}\sum_{i=1}^{|\mathbf{c}|} \log p(c_i | c_{<i})$$

Lang	ISO	% Infl	LM	MTL	Δ
Bulgarian	BG	39%	1.890	1.887	0.003
Catalan	CA	31%	1.653	1.599	0.054
Czech	CS	43%	2.045	1.832	0.213
Danish	DA	30%	2.152	2.135	0.017
German	DE	33%	1.917	1.881	0.036
English	EN	15%	2.183	2.173	0.010
Spanish	ES	28%	1.801	1.763	0.038
Farsi	FA	27%	2.213	2.205	0.008
French	FR	32%	1.751	1.712	0.039
Hindi	HI	28%	1.819	1.773	0.046
Croatian	HR	49%	1.866	1.841	0.025
Italian	IT	36%	1.595	1.554	0.041
Latvian	LV	47%	2.243	2.217	0.026
Dutch	NL	19%	1.989	1.972	0.017
Polish	PL	42%	2.218	2.154	0.064
Portuguese	PT	31%	1.787	1.785	0.002
Romanian	RO	42%	1.840	1.798	0.042
Russian	RU	42%	1.993	1.824	0.169
Slovak	SK	45%	2.705	2.686	0.019
Ukranian	UK	40%	2.359	2.338	0.021
Estonian	ET	49%	2.089	1.993	0.096
Finnish	FI	55%	1.981	1.921	0.060
Arabic	AR	86%	1.724	1.708	0.016
Hebrew	HE	42%	2.293	2.282	0.011

Language Modeling: Fully Supervised Setting

CLMs trained with Universal Dependencies for both LM, morphology supervision

MTL improves over LM baseline on all 24 languages

See biggest gains in BPC on RU and CS

$$\mathrm{BPC}(\mathbf{c}) = -\frac{1}{|\mathbf{c}|}\sum_{i=1}^{|\mathbf{c}|} \log p(c_i|c_{<i})$$

Lang	ISO	% Infl	LM	MTL	Δ
Bulgarian	BG	39%	1.890	1.887	0.003
Catalan	CA	31%	1.653	1.599	0.054
Czech	CS	43%	2.045	1.832	0.213
Danish	DA	30%	2.152	2.135	0.017
German	DE	33%	1.917	1.881	0.036
English	EN	15%	2.183	2.173	0.010
Spanish	ES	28%	1.801	1.763	0.038
Farsi	FA	27%	2.213	2.205	0.008
French	FR	32%	1.751	1.712	0.039
Hindi	HI	28%	1.819	1.773	0.046
Croatian	HR	49%	1.866	1.841	0.025
Italian	IT	36%	1.595	1.554	0.041
Latvian	LV	47%	2.243	2.217	0.026
Dutch	NL	19%	1.989	1.972	0.017
Polish	PL	42%	2.218	2.154	0.064
Portuguese	PT	31%	1.787	1.785	0.002
Romanian	RO	42%	1.840	1.798	0.042
Russian	RU	42%	1.993	1.824	0.169
Slovak	SK	45%	2.705	2.686	0.019
Ukranian	UK	40%	2.359	2.338	0.021
Estonian	ET	49%	2.089	1.993	0.096
Finnish	FI	55%	1.981	1.921	0.060
Arabic	AR	86%	1.724	1.708	0.016
Hebrew	HE	42%	2.293	2.282	0.011

Typology 101

Fusional: one form of a morpheme can simultaneously encode several meanings (e.g., English, Russian, Spanish)

Lang	ISO	% Infl	LM	MTL	Δ
Bulgarian	BG	39%	1.890	1.887	0.003
Catalan	CA	31%	1.653	1.599	0.054
Czech	CS	43%	2.045	1.832	0.213
Danish	DA	30%	2.152	2.135	0.017
German	DE	33%	1.917	1.881	0.036
English	EN	15%	2.183	2.173	0.010
Spanish	ES	28%	1.801	1.763	0.038
Farsi	FA	27%	2.213	2.205	0.008
French	FR	32%	1.751	1.712	0.039
Hindi	HI	28%	1.819	1.773	0.046
Croatian	HR	49%	1.866	1.841	0.025
Italian	IT	36%	1.595	1.554	0.041
Latvian	LV	47%	2.243	2.217	0.026
Dutch	NL	19%	1.989	1.972	0.017
Polish	PL	42%	2.218	2.154	0.064
Portuguese	PT	31%	1.787	1.785	0.002
Romanian	RO	42%	1.840	1.798	0.042
Russian	RU	42%	1.993	1.824	0.169
Slovak	SK	45%	2.705	2.686	0.019
Ukranian	UK	40%	2.359	2.338	0.021
Estonian	ET	49%	2.089	1.993	0.096
Finnish	FI	55%	1.981	1.921	0.060
Arabic	AR	86%	1.724	1.708	0.016
Hebrew	HE	42%	2.293	2.282	0.011

Typology 101

Agglutinative: words are made up of a linear sequence of distinct morphemes and each component of meaning is represented by its own morpheme.

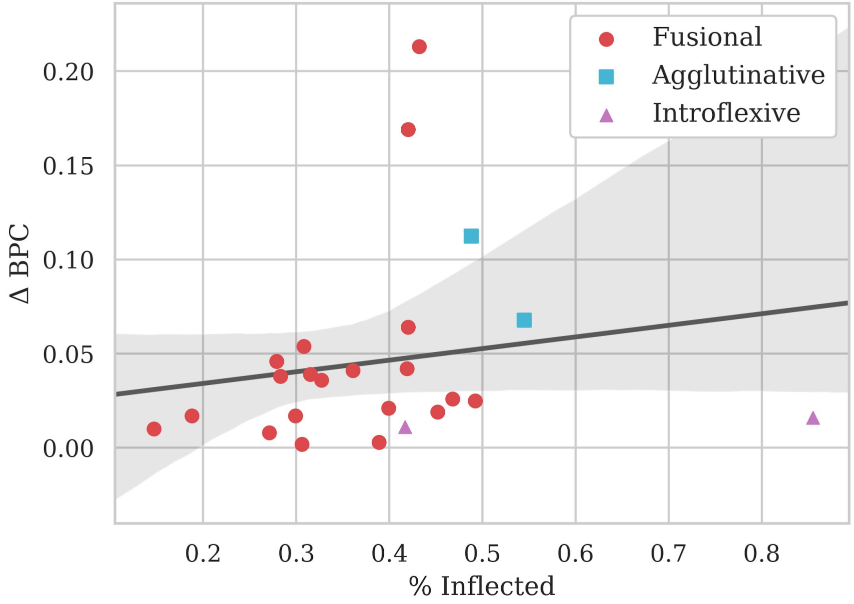
Lang	ISO	% Infl	LM	MTL	Δ
Bulgarian	BG	39%	1.890	1.887	0.003
Catalan	CA	31%	1.653	1.599	0.054
Czech	CS	43%	2.045	1.832	0.213
Danish	DA	30%	2.152	2.135	0.017
German	DE	33%	1.917	1.881	0.036
English	EN	15%	2.183	2.173	0.010
Spanish	ES	28%	1.801	1.763	0.038
Farsi	FA	27%	2.213	2.205	0.008
French	FR	32%	1.751	1.712	0.039
Hindi	HI	28%	1.819	1.773	0.046
Croatian	HR	49%	1.866	1.841	0.025
Italian	IT	36%	1.595	1.554	0.041
Latvian	LV	47%	2.243	2.217	0.026
Dutch	NL	19%	1.989	1.972	0.017
Polish	PL	42%	2.218	2.154	0.064
Portuguese	PT	31%	1.787	1.785	0.002
Romanian	RO	42%	1.840	1.798	0.042
Russian	RU	42%	1.993	1.824	0.169
Slovak	SK	45%	2.705	2.686	0.019
Ukranian	UK	40%	2.359	2.338	0.021
Estonian	ET	49%	2.089	1.993	0.096
Finnish	FI	55%	1.981	1.921	0.060
Arabic	AR	86%	1.724	1.708	0.016
Hebrew	HE	42%	2.293	2.282	0.011

Typology 101

Introflexive: words are inflected into different forms through the insertion of a pattern of vowels into a consonantal root.

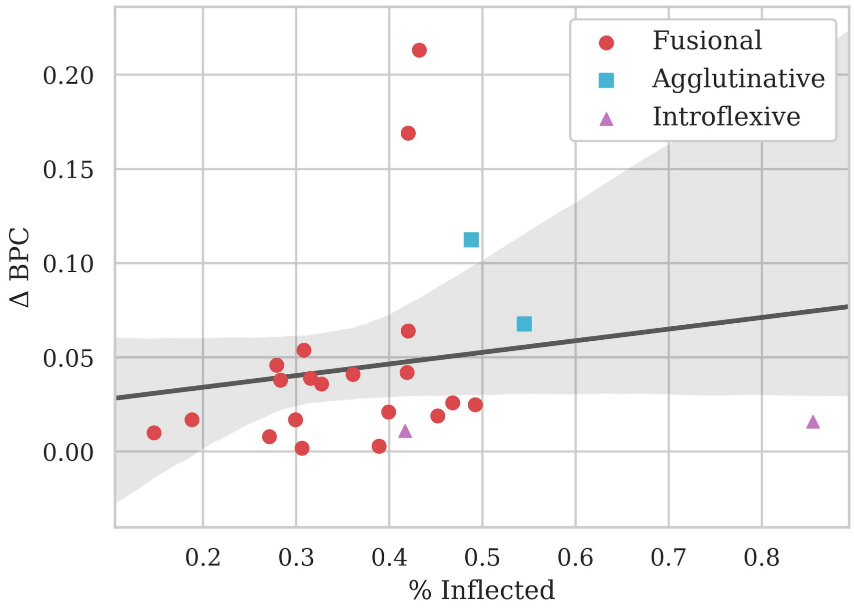
Lang	ISO	% Infl	LM	MTL	Δ
Bulgarian	BG	39%	1.890	1.887	0.003
Catalan	CA	31%	1.653	1.599	0.054
Czech	CS	43%	2.045	1.832	0.213
Danish	DA	30%	2.152	2.135	0.017
German	DE	33%	1.917	1.881	0.036
English	EN	15%	2.183	2.173	0.010
Spanish	ES	28%	1.801	1.763	0.038
Farsi	FA	27%	2.213	2.205	0.008
French	FR	32%	1.751	1.712	0.039
Hindi	HI	28%	1.819	1.773	0.046
Croatian	HR	49%	1.866	1.841	0.025
Italian	IT	36%	1.595	1.554	0.041
Latvian	LV	47%	2.243	2.217	0.026
Dutch	NL	19%	1.989	1.972	0.017
Polish	PL	42%	2.218	2.154	0.064
Portuguese	PT	31%	1.787	1.785	0.002
Romanian	RO	42%	1.840	1.798	0.042
Russian	RU	42%	1.993	1.824	0.169
Slovak	SK	45%	2.705	2.686	0.019
Ukranian	UK	40%	2.359	2.338	0.021
Estonian	ET	49%	2.089	1.993	0.096
Finnish	FI	55%	1.981	1.921	0.060
Arabic	AR	86%	1.724	1.708	0.016
Hebrew	HE	42%	2.293	2.282	0.011

Analysis of Fully Supervised MTL on UD

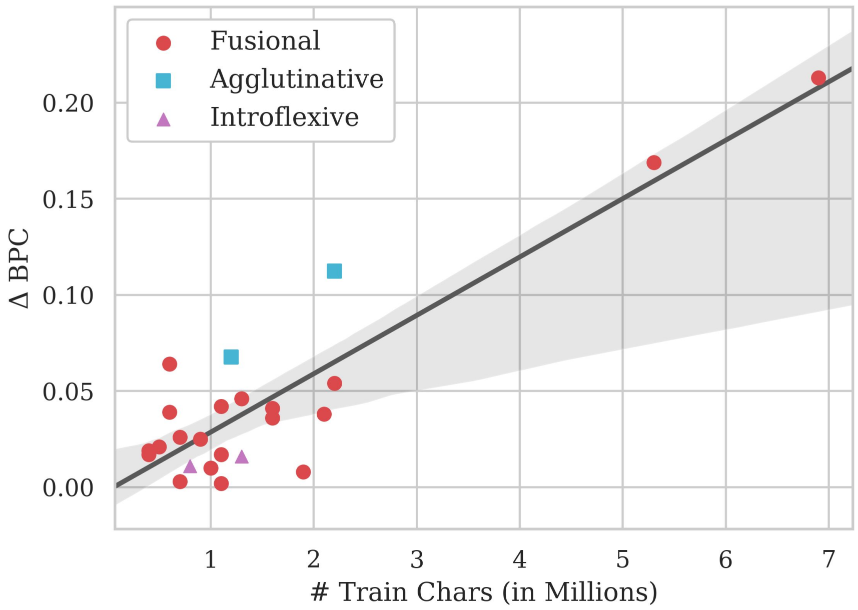


r = 0.152

Analysis of Fully Supervised MTL on UD

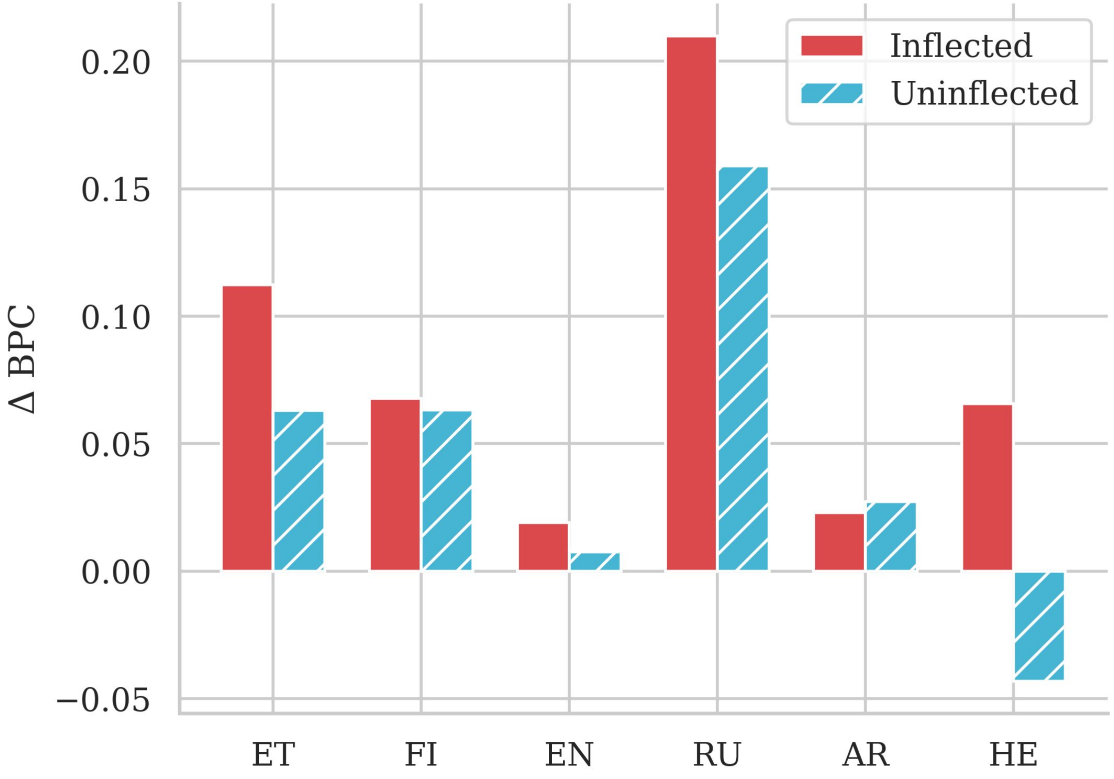


r = 0.152

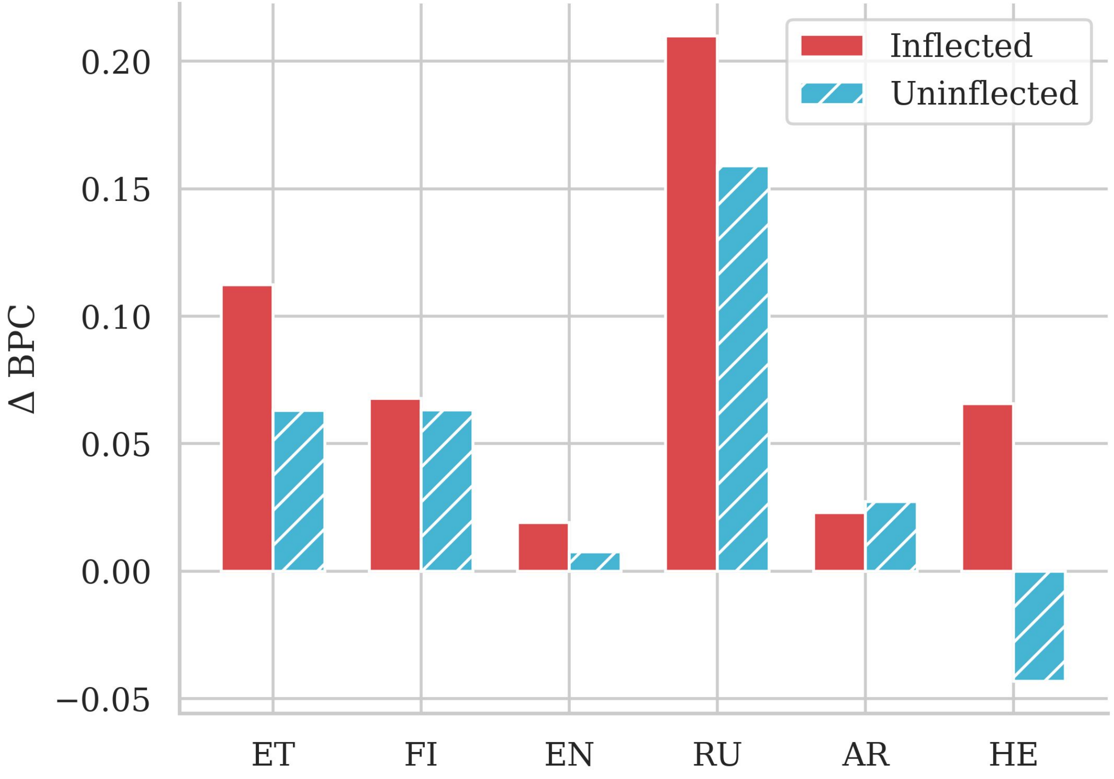


r = 0.931

BPC Improvement on Inflected vs. Uninflected Forms



BPC Improvement on Inflected vs. Uninflected Forms



Better BPC gains on inflected forms for 16 out of 24 languages

Across languages, BPC on inflected forms is **31%** better than on uninflected forms

	CS		DE		EN		ES		FI		FR		RU	
	dev	test	dev	test	dev	test	dev	test	dev	test	dev	test	dev	test
HCLM	2.010	1.984	1.605	1.588	1.591	1.538	1.548	1.498	1.754	1.711	1.499	1.467	1.777	1.761
LM	2.013	1.972	1.557	1.538	1.543	1.488	1.571	1.505	1.725	1.699	1.357	1.305	1.745	1.724
MTL	**1.938**	**1.900**	**1.249**	**1.241**	**1.313**	**1.256**	**1.260**	**1.196**	**1.698**	**1.669**	**1.211**	**1.167**	**1.645**	**1.619**
Δ	0.075	0.072	0.308	0.297	0.230	0.232	0.311	0.309	0.027	0.030	0.146	0.138	0.100	0.105

Language Modeling: Distantly Supervised Setting

Models trained with Multilingual Wikipedia Corpus (MWC) for LM supervision, UD annotations for morphology supervision

	CS		DE		EN		ES		FI		FR		RU	
	dev	test	dev	test	dev	test	dev	test	dev	test	dev	test	dev	test
HCLM	2.010	1.984	1.605	1.588	1.591	1.538	1.548	1.498	1.754	1.711	1.499	1.467	1.777	1.761
LM	2.013	1.972	1.557	1.538	1.543	1.488	1.571	1.505	1.725	1.699	1.357	1.305	1.745	1.724
MTL	**1.938**	**1.900**	**1.249**	**1.241**	**1.313**	**1.256**	**1.260**	**1.196**	**1.698**	**1.669**	**1.211**	**1.167**	**1.645**	**1.619**
Δ	0.075	0.072	0.308	0.297	0.230	0.232	0.311	0.309	0.027	0.030	0.146	0.138	0.100	0.105

Language Modeling: Distantly Supervised Setting

Models trained with Multilingual Wikipedia Corpus (MWC) for LM supervision, UD annotations for morphology supervision

MTL improves over LM baseline and a more complex architecture from Kawakami et al. (2017), HCLMcache

Kazuya Kawakami et al.. Learning to create and reuse words in open-vocabulary neural language modeling. In *ACL*, 2017.

	CS		DE		EN		ES		FI		FR		RU	
	dev	test	dev	test	dev	test	dev	test	dev	test	dev	test	dev	test
HCLM	2.010	1.984	1.605	1.588	1.591	1.538	1.548	1.498	1.754	1.711	1.499	1.467	1.777	1.761
LM	2.013	1.972	1.557	1.538	1.543	1.488	1.571	1.505	1.725	1.699	1.357	1.305	1.745	1.724
MTL	**1.938**	**1.900**	**1.249**	**1.241**	**1.313**	**1.256**	**1.260**	**1.196**	**1.698**	**1.669**	**1.211**	**1.167**	**1.645**	**1.619**
Δ	0.075	0.072	0.308	$\boxed{0.297}$	0.230	$\boxed{0.232}$	0.311	$\boxed{0.309}$	0.027	0.030	0.146	0.138	0.100	0.105

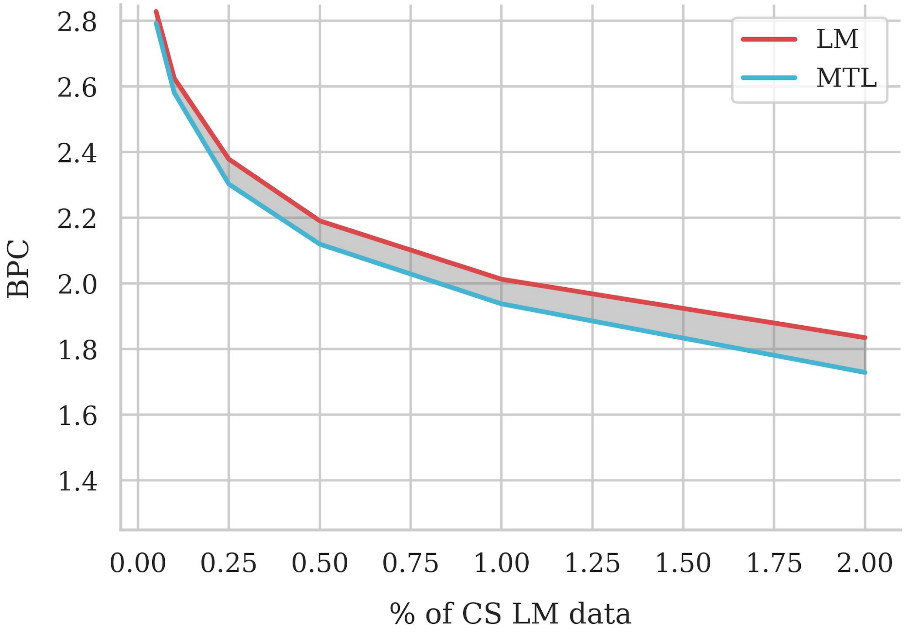
Language Modeling: Distantly Supervised Setting

Models trained with Multilingual Wikipedia Corpus (MWC) for LM supervision, UD annotations for morphology supervision

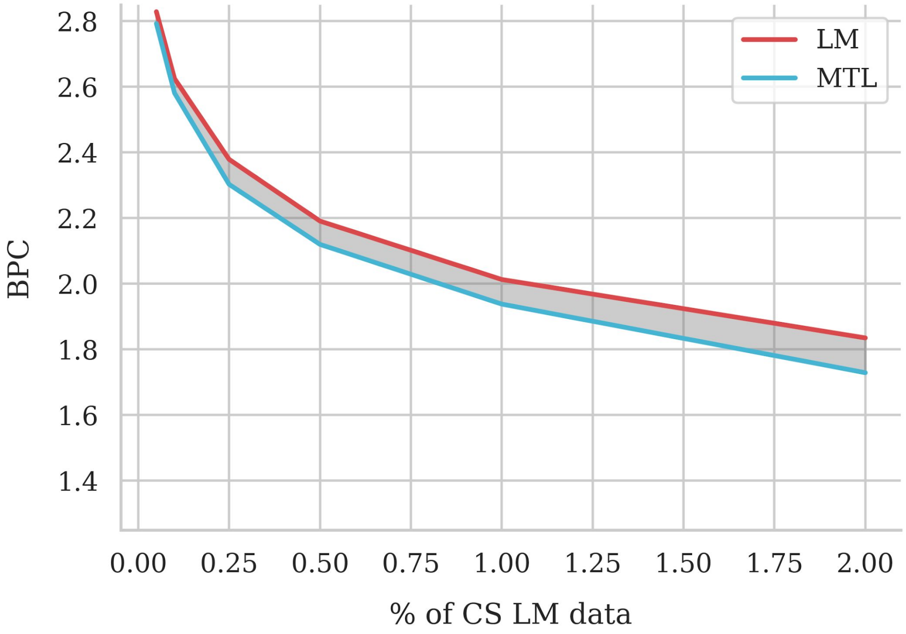
MTL improves over LM baseline more complex architecture from Kawakami et al. (2017), HCLMcache

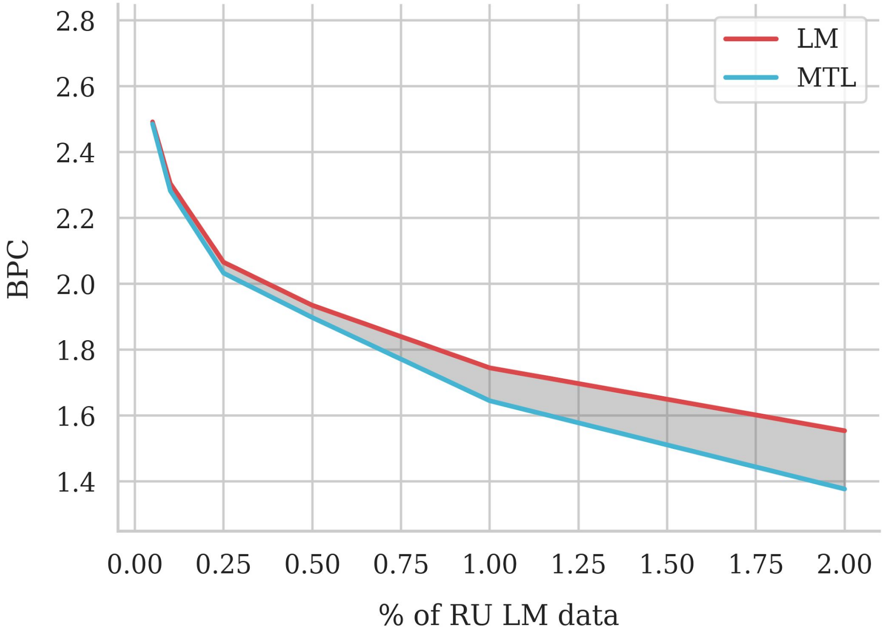
Better BPC gains on languages with more LM data (DE, EN, ES)

How does the amount of LM data affect BPC?

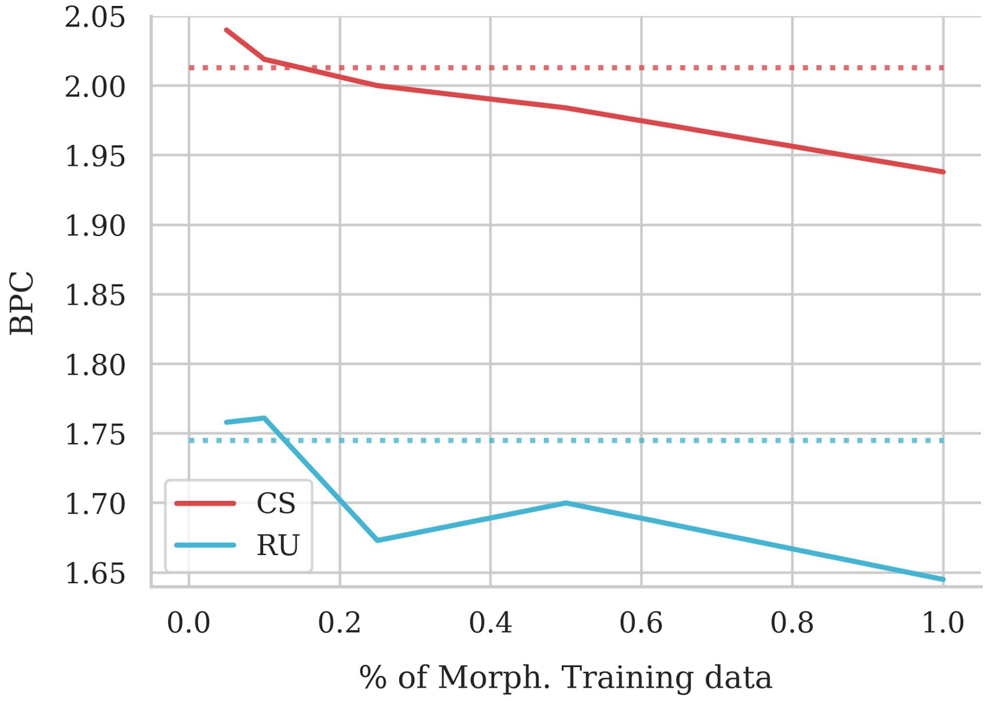


How does the amount of LM data affect BPC?





How does the amount of labeled morphology data affect BPC?



Cross-Lingual Transfer

LM data	Morph. data	BPC
SK	None	2.806
	SK	2.779
	CS	2.752
	CS+SK	2.777
CS+SK	None	2.668
	CS+SK	**2.446**
UK	None	2.369
	UK	2.348
	RU	2.348
	RU+UK	2.351
RU+UK	None	2.495
	RU+UK	**2.316**

Cross-Lingual Transfer

Czech (CS) -> Slovak (SK)
6.9M chars 0.4M chars

Russian (RU) -> Ukrainian (UK)
5.3M chars 0.5M chars

LM data	Morph. data	BPC
SK	None	2.806
	SK	2.779
	CS	2.752
	CS+SK	2.777
CS+SK	None	2.668
	CS+SK	**2.446**
UK	None	2.369
	UK	2.348
	RU	2.348
	RU+UK	2.351
RU+UK	None	2.495
	RU+UK	**2.316**

Cross-Lingual Transfer

Czech (CS) -> Slovak (SK)

6.9M chars 0.4M chars

Russian (RU) -> Ukrainian (UK)

5.3M chars 0.5M chars

Best BPC on low-resource language from sharing LM and morph data

LM data	Morph. data	BPC
SK	None	2.806
	SK	2.779
	CS	2.752
	CS+SK	2.777
CS+SK	None	2.668
	CS+SK	**2.446**
UK	None	2.369
	UK	2.348
	RU	2.348
	RU+UK	2.351
RU+UK	None	2.495
	RU+UK	**2.316**

Cross-Lingual Transfer

Czech (CS) -> Slovak (SK)
6.9M chars 0.4M chars

Russian (RU) -> Ukrainian (UK)
5.3M chars 0.5M chars

Best BPC on low-resource language from sharing LM and morph data

CS+SK MTL improves by 0.333 BPC over SK MTL

RU+UK MTL improves by 0.032 BPC over UK MTL

LM data	Morph. data	BPC
SK	None	2.806
	SK	2.779
	CS	2.752
	CS+SK	2.777
CS+SK	None	2.668
	CS+SK	**2.446**
UK	None	2.369
	UK	2.348
	RU	2.348
	RU+UK	2.351
RU+UK	None	2.495
	RU+UK	**2.316**

Related Work

Modifying architecture for morphologically-rich languages

Kazuya Kawakami et al.. Learning to create and reuse words in open-vocabulary neural language modeling. In *ACL*, 2017.

Daniela Gerz et al. Language modeling for morphologically rich languages: Character-aware modeling for word-level prediction. *TACL*, 2018.

Sebastian J. Mielke and Jason Eisner. Spell once, summon anywhere: A two-level open-vocabulary language model. In *AAAI*, 2019

Related Work

Adding morphology as input to the model

Clara Vania and Adam Lopez. From characters to words to in between: Do we capture morphology? In *ACL*, 2017.

Jan Botha and Phil Blunsom. Compositional morphology for word representations and language modeling. In *ICML*, 2014

Austin Matthews et al., Using morphological knowledge in open-vocabulary language models. In *NAACL*, 2018.

Related Work

<u>Multitasking morphology into decoder of NMT system</u>:

Fahim Dalvi et al., Understanding and Improving morphological learning in the neural machine translation decoder. In *IJCNLP*, 2017.

In Conclusion…

(1) Multitasking morphology with character LMs improves performance across 20+ languages

In Conclusion...

(1) Multitasking morphology with character LMs improves performance across 20+ languages

(2) BPC improves when morphology and LM datasets are disjoint -> cheap way to improve models on existing datasets

In Conclusion...

(1) Multitasking morphology with character LMs improves performance across 20+ languages

(2) BPC improves when morphology and LM datasets are disjoint -> cheap way to improve models on existing datasets

(3) BPC improves more on inflected forms than uninflected forms

In Conclusion...

(1) Multitasking morphology with character LMs improves performance across 20+ languages

(2) BPC improves when morphology and LM datasets are disjoint -> cheap way to improve models on existing datasets

(3) BPC improves more on inflected forms than uninflected forms

(4) Increasing the amount of raw text available to the model does not reduce gains in BPC

In Conclusion...

(1) Multitasking morphology with character LMs improves performance across 20+ languages

(2) BPC improves when morphology and LM datasets are disjoint -> cheap way to improve models on existing datasets

(3) BPC improves more on inflected forms than uninflected forms

(4) Increasing the amount of raw text available to the model does not reduce gains in BPC **-- in fact, it improves it!**

In Conclusion...

(1) Multitasking morphology with character LMs improves performance across 20+ languages

(2) BPC improves when morphology and LM datasets are disjoint -> cheap way to improve models on existing datasets

(3) BPC improves more on inflected forms than uninflected forms

(4) Increasing the amount of raw text available to the model does not reduce gains in BPC **-- in fact, it improves it!**

(5) Morphology annotations can be shared across related languages to improve LM in a low-resource setting

Thank you!