Identifying negative language transfer in learner writing: Using syntactic information to model structural differences

Leticia Farias Wanderley
May 25, 2021

EdTeKLA

What is negative language transfer?

- A second language acquisition phenomenon
- Language learners reuse their native languages' grammar rules when communicating in a second language
- When the reused rules are different from second language rules, negative language transfer occurs

Negative language transfer example

Negative language transfer example

“The idea of international art festival was great.”

Negative language transfer example

“The idea of international art festival was great.”

This sentence is missing an article

Negative language transfer example

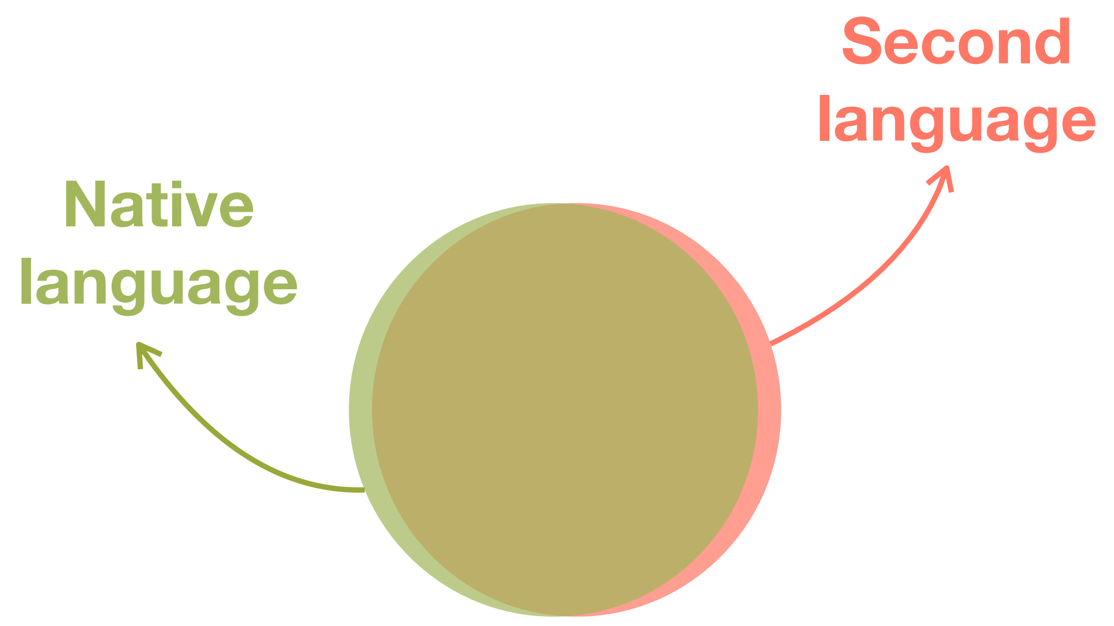
“The idea of **an** international art festival was great.”

Intuition

Second language
learning begins

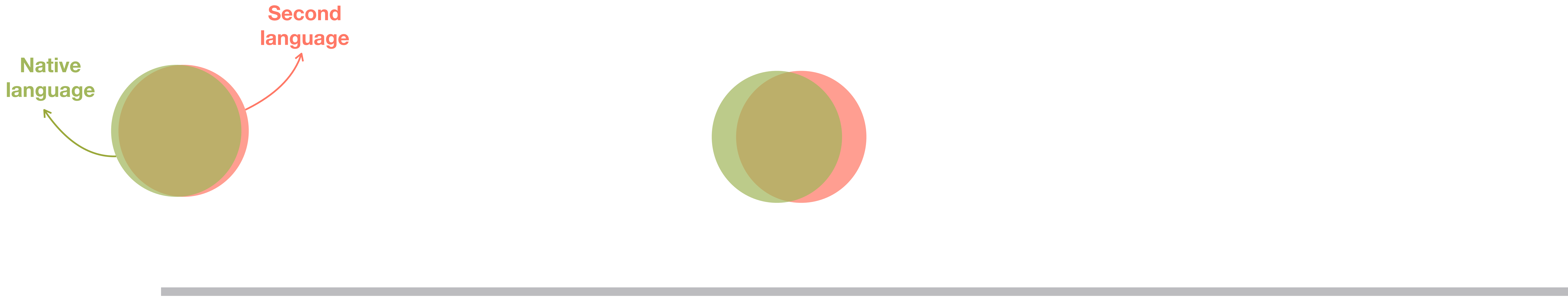
Intuition

Intuition



Second language
learning begins

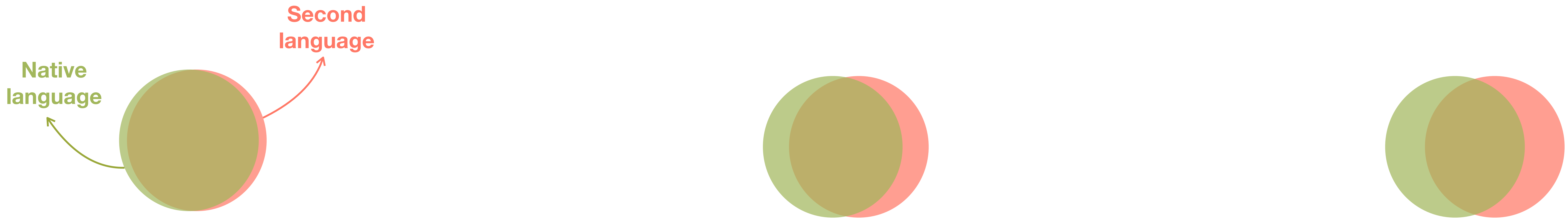
Intuition



Second language
learning begins

Language learner continues
to acquire the second language

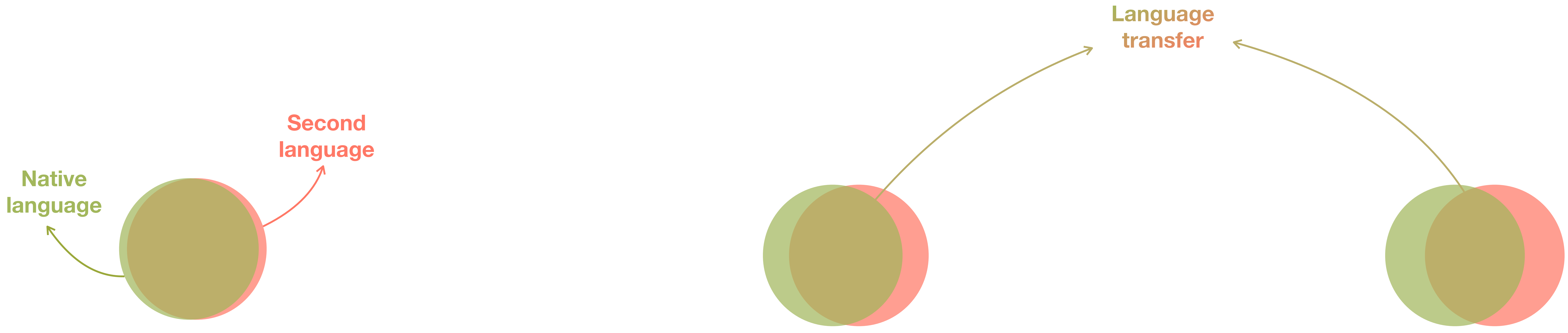
Intuition



Second language
learning begins

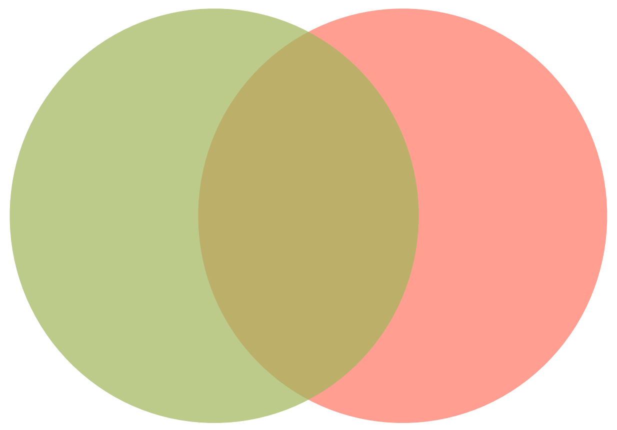
Language learner continues
to acquire the second language

Intuition



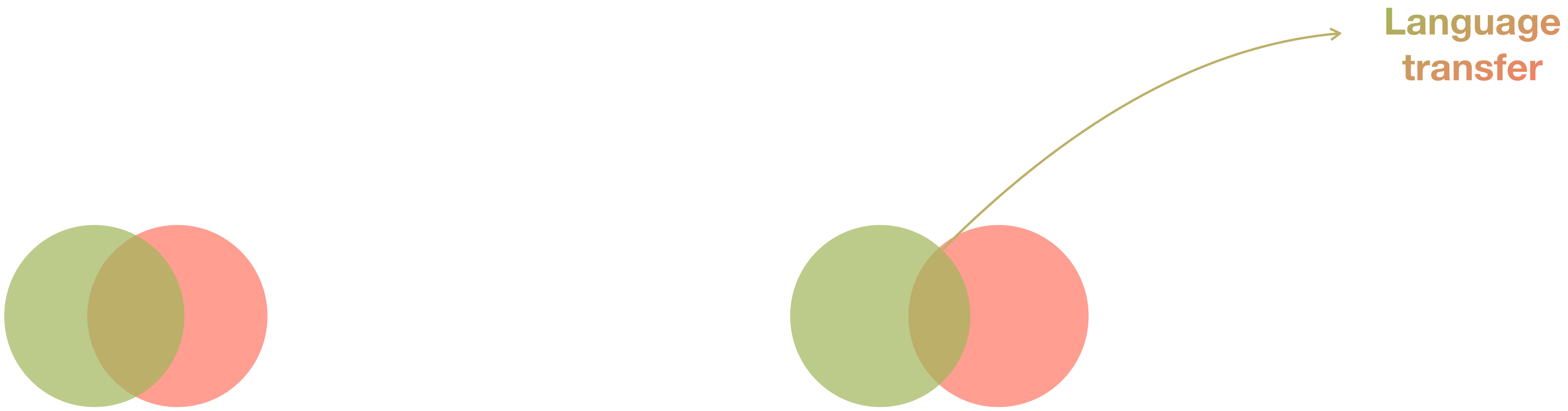
Second language
learning begins

Language learner continues
to acquire the second language



Learner becomes proficient
in the second language

Learner becomes proficient in the second language



Learner becomes proficient
in the second language

Interlanguage theory

- When learning a second language, learners maintain a linguistic system that evolves as they acquire the language

- Many processes influence the development of an interlanguage, one of them is language transfer

Applications in natural language processing

- Native language identification
  - Error patterns can be used as native language predictors

Applications in natural language processing

- Native language identification
  - Error patterns can be used as native language predictors
- Grammatical error correction
  - L1-specific data can improve GEC system performance
  - Contrastive feedback for missing preposition errors

Applications in natural language processing

- Native language identification
  - Error patterns can be used as native language predictors
- Grammatical error correction
  - L1-specific data can improve GEC system performance
  - Contrastive feedback for missing preposition errors
- There hasn't been a more general approach to negative language transfer detection

Language learners could benefit from being more aware of this phenomenon

Metalinguistic feedback

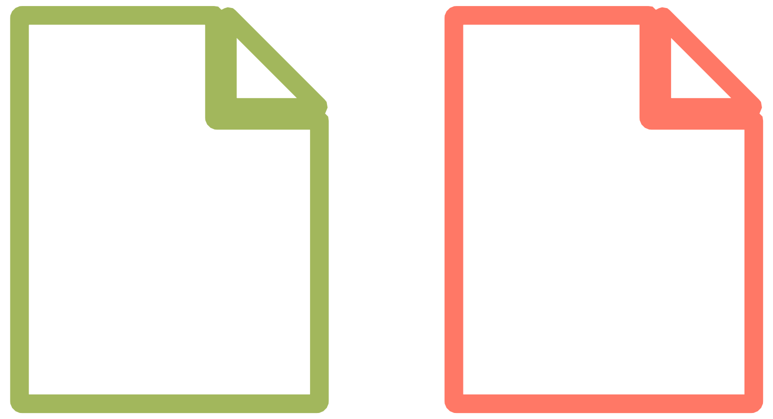
- Language learners benefit from error feedback
- Explanations about error causes can help learners understand why they made a mistake
- **And prevent them from making the mistake again**

First steps

- Develop a method to identify when language learner errors are related to language transfer
- Evaluate the method on errors made by Chinese native speakers who are learning English

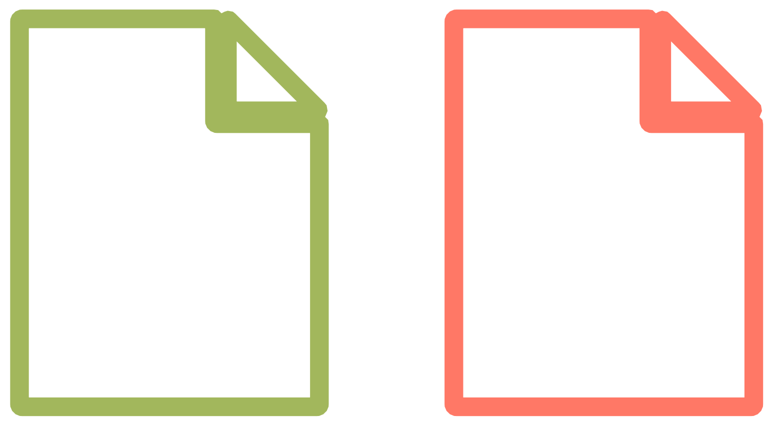
Create models that can differentiate between English and Chinese language structures. Then, use those models to identify Chinese patterns in learner errors

Methodology overview



Parallel textual data in Chinese and English

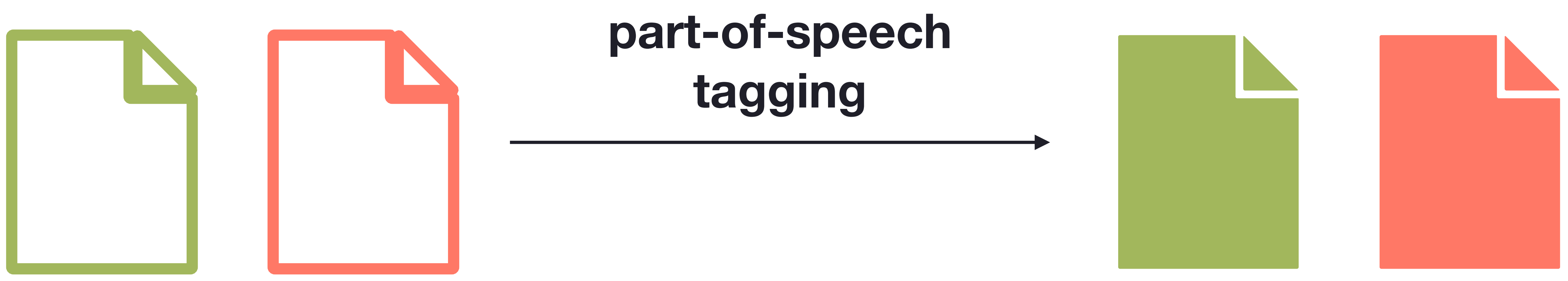
Methodology overview



Parallel textual data in Chinese and English

part-of-speech tagging →

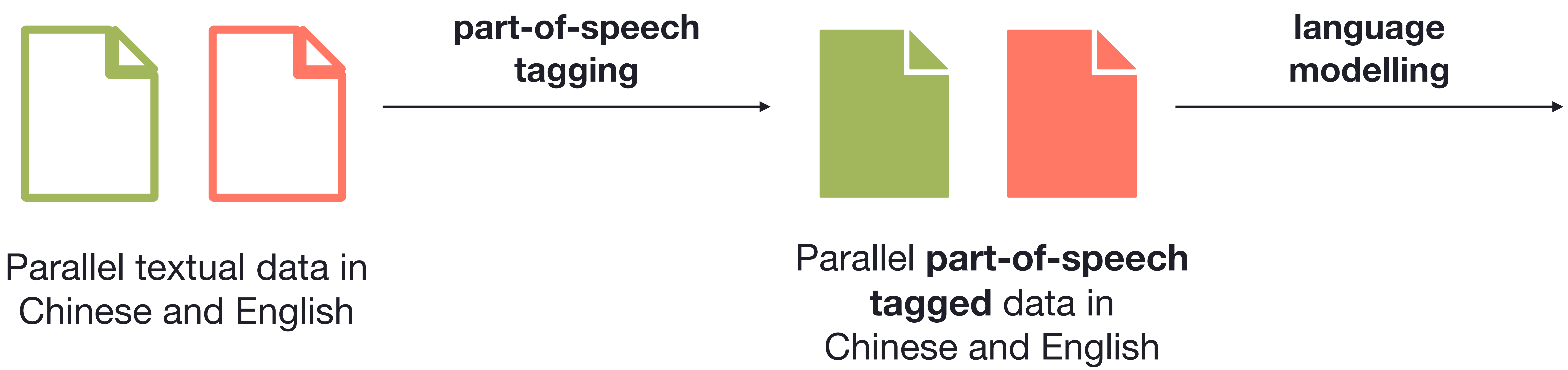
Methodology overview



Parallel textual data in Chinese and English

Parallel **part-of-speech tagged** data in Chinese and English

Methodology overview



Methodology overview

Methodology overview

Part-of-speech tag sequence extracted from a learner error → **Part-of-speech language models**

Methodology overview

Part-of-speech tag sequence extracted from a learner error → **Part-of-speech language models**

Methodology overview

Part-of-speech tag sequence extracted from a learner error

Part-of-speech language models

It **is** related to negative language transfer

It is **not** related to negative language transfer

Training data

- Manually translated parallel datasets
- The datasets were aligned at the sentence level
- 150K POS tag sequences extracted from English data and 150K extracted from Chinese data

Training data

Global Voices dataset	138 582
WMT19 - Machine Translation of News	11 960
Total	**150 542**

Test data

- Learner errors extracted from the First Certificate in English dataset
- All the errors were annotated with information about their connection to negative language transfer
- More than 3000 learner errors were annotated

Leticia Farias Wanderley, Nicole Zhao, and Carrie Demmans Epp. Negative language transfer in learner English: A new dataset. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 3129–3142, Online, June 2021. Association for Computational Linguistics.

Negative language transfer annotation

	Negative transfer?	Why?
The idea of international art festival was great.		

Leticia Farias Wanderley, Nicole Zhao, and Carrie Demmans Epp. Negative language transfer in learner English: A new dataset. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 3129–3142, Online, June 2021. Association for Computational Linguistics.

Negative language transfer annotation

	Negative transfer?	Why?
The idea of international art festival was great.	✅	

Leticia Farias Wanderley, Nicole Zhao, and Carrie Demmans Epp. Negative language transfer in learner English: A new dataset. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 3129–3142, Online, June 2021. Association for Computational Linguistics.

Negative language transfer annotation

	Negative transfer?	Why?
The idea of international art festival was great.	✅	Chinese has no articles and doesn't use classifiers in this situation

Leticia Farias Wanderley, Nicole Zhao, and Carrie Demmans Epp. Negative language transfer in learner English: A new dataset. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 3129–3142, Online, June 2021. Association for Computational Linguistics.

Structural learner errors

Errors that can be represented by POS tag sequences

Negative language transfer errors	1457
Not negative language transfer errors	914
Total	**2371**

How much of the errors' contexts should be included in the test sequences?

"I only can travel in July"

How much of the errors' contexts should be included in the test sequences?

"I only can travel in July"

How much of the errors' contexts should be included in the test sequences?

"I <u>only</u> <u>can</u> travel in July"

How much of the errors' contexts should be included in the test sequences?

"I **can only** travel in July"

How much of the errors' contexts should be included in the test sequences?

"I only can travel in July"

How much of the errors' contexts should be included in the test sequences?

"I only can travel in July"

PRON ADV VERB VERB ADP NOUN

How much of the errors' contexts should be included in the test sequences?

"I <u>only</u> <u>can</u> travel in July"

ADV VERB

Test sequences

Padded error span “I only can travel in July”

Test sequences

Padded error span “<u>I</u> <u>only</u> <u>can</u> <u>travel</u> in July”

PRON ADV VERB VERB

Test sequences

Padded error span	"<u>I</u> <u>only</u> <u>can</u> <u>travel</u> in July" PRON ADV VERB VERB
Error + unigram span	"I <u>only</u> <u>can</u> <u>travel</u> in July"

Test sequences

Padded error span “<u>I</u> <u>only</u> <u>can</u> <u>travel</u> in July”

PRON ADV VERB VERB

Error + unigram span “I <u>only</u> <u>can</u> <u>travel</u> in July”

ADV VERB VERB

Test sequences

Padded error span "I only can travel in July"

PRON ADV VERB VERB

Error + unigram span "I only can travel in July"

ADV VERB VERB

Error + bigram span "I only can travel in July"

Test sequences

Padded error span “I only can travel in July”
PRON ADV VERB VERB

Error + unigram span “I only can travel in July”
ADV VERB VERB

Error + bigram span “I only can travel in July”
ADV VERB VERB ADP

Baseline language modelling approach

N-gram baseline

- Used the n-gram language model implementation from KenLM
- One n-gram model was trained with POS tag sequences extracted from English text, and the other with POS tag sequences extracted from text in Chinese
- Each model analysed sequences of 5 POS tags at a time

N-gram baseline hyperparameter tuning

- Five different n-gram lengths were analysed, from 2 to 6
- In the tuning process, models were trained on 80% of the training dataset and their accuracy was evaluated on the remaining 20% of the training data
- The best performing models (n = 5) achieved an accuracy of 96.94% on the evaluation set

N-gram baseline training procedure

- Two n-gram language models were trained
- One was trained with all POS tag sequences extracted from English sentences and the other was trained on all the POS tag sequences extracted from Chinese sentences
- Each model learnt a distribution over POS tag sequences from the training data

N-gram baseline testing procedure

Part-of-speech tag
sequence extracted
from a learner error

N-gram baseline testing procedure

Part-of-speech tag sequence extracted from a learner error

Chinese n-gram model

English n-gram model

N-gram baseline testing procedure

Part-of-speech tag sequence extracted from a learner error

Chinese n-gram model → Chinese probability

English n-gram model → English probability

N-gram baseline testing procedure

Part-of-speech tag sequence extracted from a learner error

Chinese n-gram model → Chinese probability

English n-gram model → English probability

If Chinese probability > English probability, classify learner error as negative language transfer

N-gram baseline results

Span	P	R	F1
Padded error	**0.68**	0.32	0.43
Error + unigram	0.64	**0.34**	**0.45**
Error + bigram	0.66	0.27	0.38

Limitation

- The English model and the Chinese model are independent from one another
- Each model's output only represents the likelihood of the POS tag sequence belonging to the language structure it models

Limitation: independent models

Part-of-speech tag sequence extracted from a learner error

Chinese n-gram model → **Chinese probability**

English n-gram model → **English probability**

If Chinese probability > English probability, classify learner error as negative language transfer

RNN language modelling approach

RNN approach

- Used the RNN implementation from PyTorch
- One single network learnt to differentiate between Chinese and English structures from the training data
- The model was trained for 10 epochs with Adam optimization. It had 16 hidden units, learning rate of 0.0001, mini batch size = 1, and negative log likelihood as its loss function

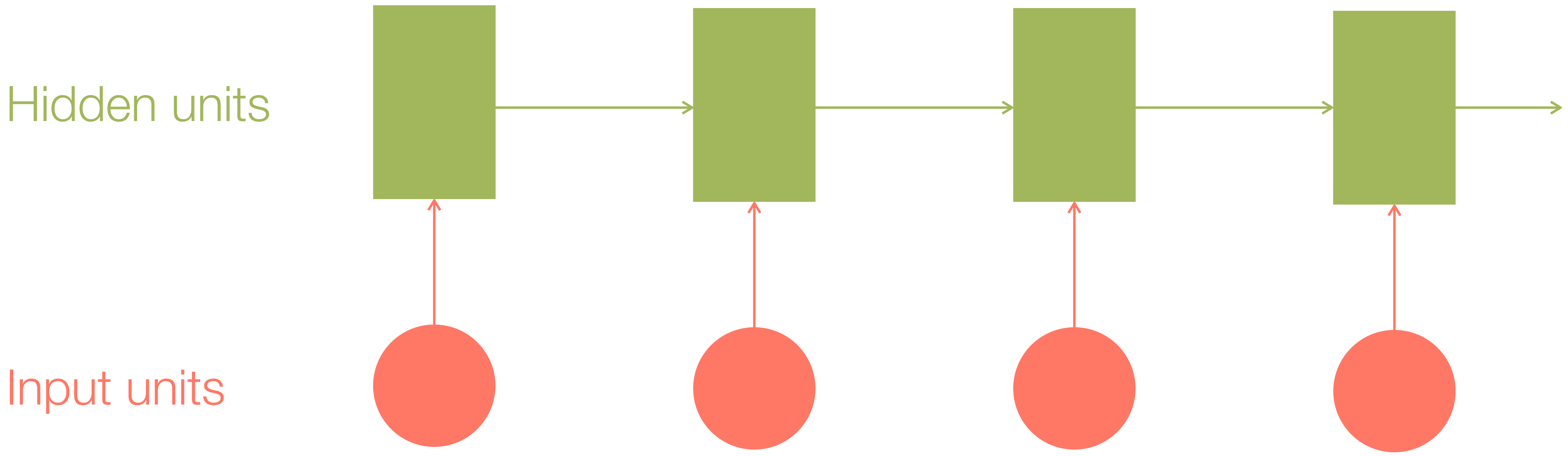
RNN approach hyperparameter tuning

- The number of hidden units, learning rate, mini batch size, and loss function were the hyperparameters tuned
- Hyperparameter combination performances were defined as their language source prediction accuracy
- The best performing model achieved an accuracy of 95.16% on the evaluation set

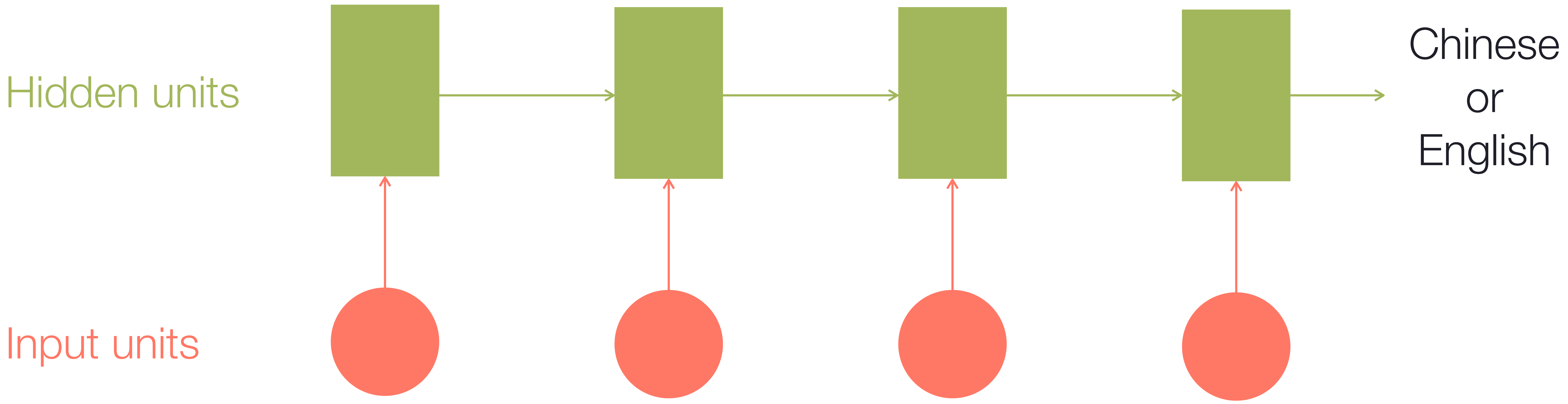
RNN approach training procedure

- The RNN model was trained with all the POS tag sequences extracted from Chinese and English sentences
- POS tags were represented as one-hot encoding vectors
- The RNN model learnt to predict a source language from a POS tag sequence

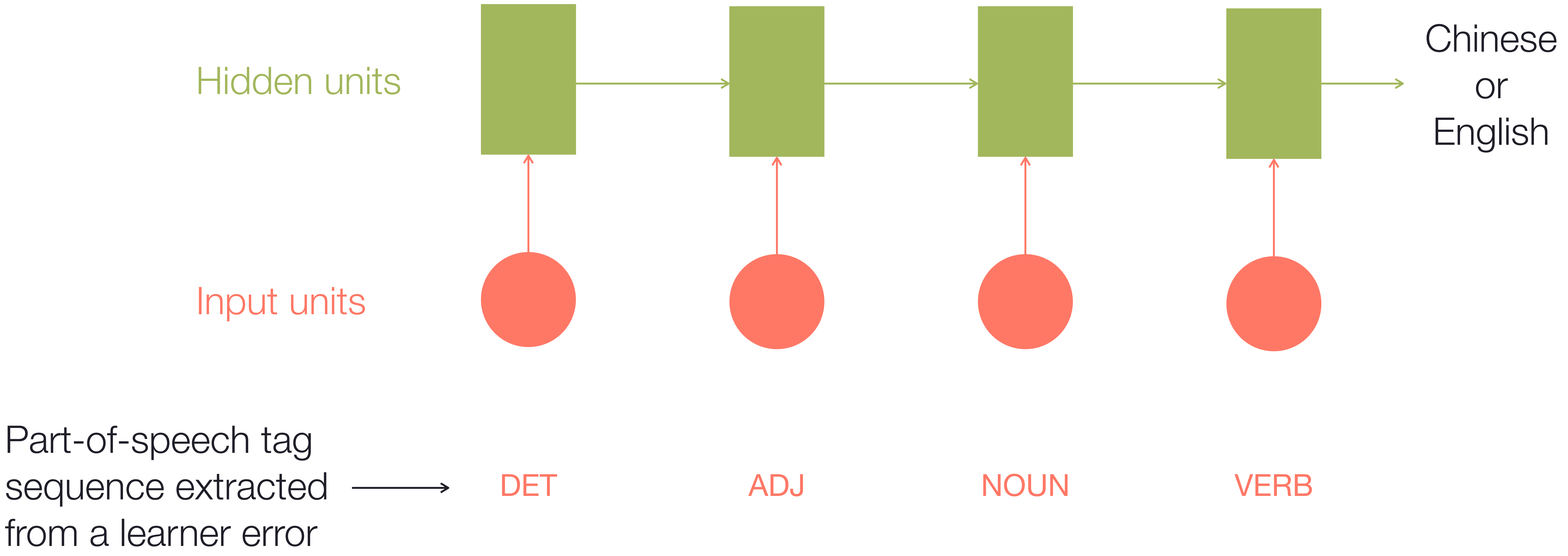
RNN approach testing procedure



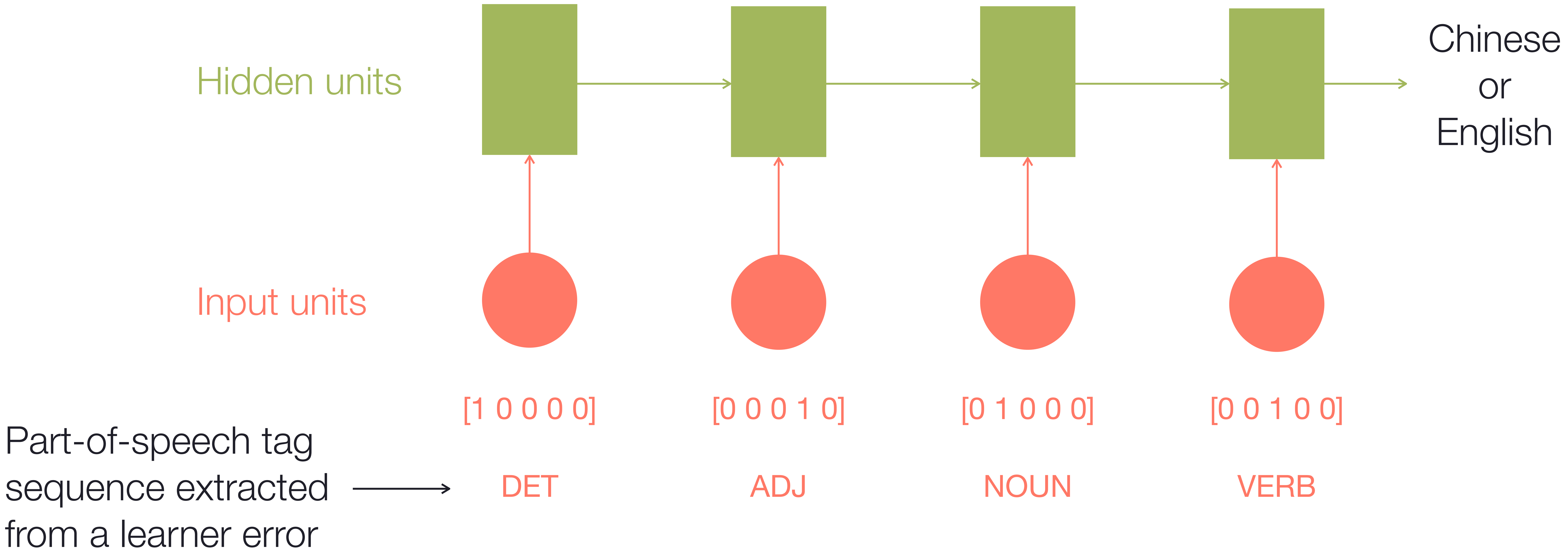
RNN approach testing procedure



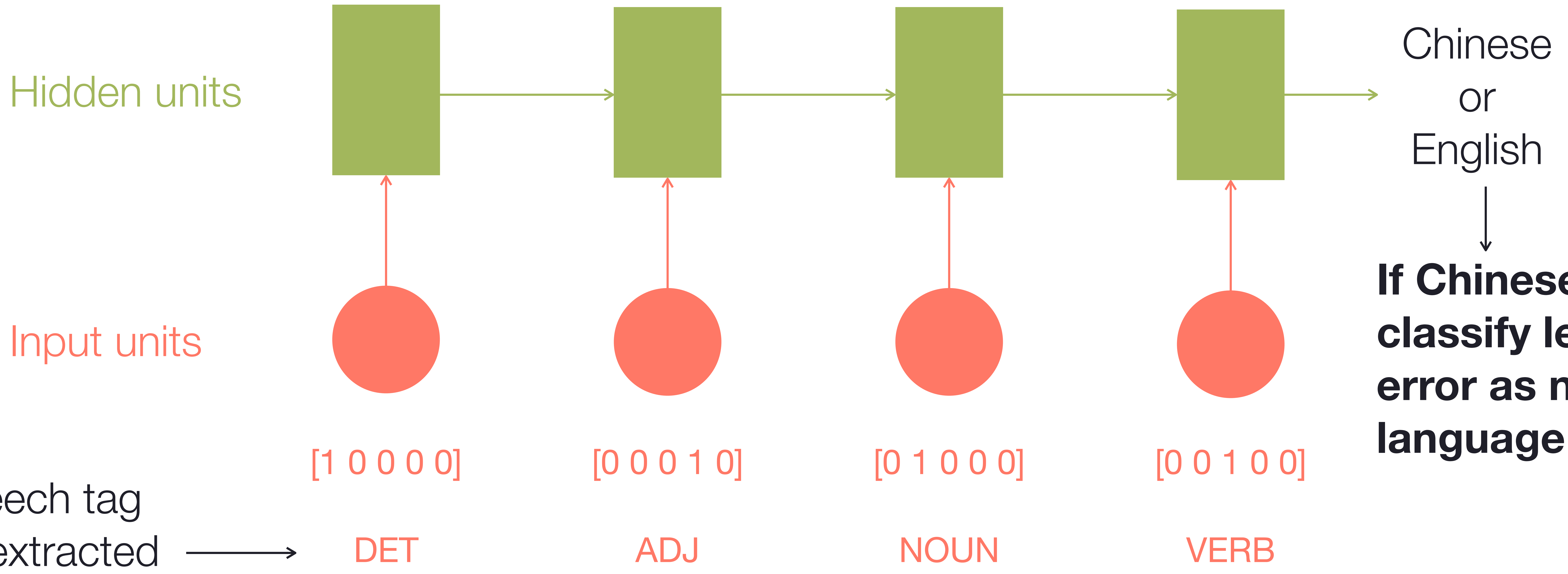
RNN approach testing procedure



RNN approach testing procedure



RNN approach testing procedure



Part-of-speech tag sequence extracted from a learner error

If Chinese, classify learner error as negative language transfer

RNN approach results

Span	**P**	**R**	**F1**
Padded error	0.69	0.34	0.46
Error + unigram	0.67	**0.41**	**0.51**
Error + bigram	**0.70**	0.35	0.46

Results

Approach	Span	P	R	F1
N-gram baseline	Padded error	0.68	0.32	0.43
	Error + unigram			
	Error + bigram			
RNN	Padded error	0.69	0.34	0.46
	Error + unigram			
	Error + bigram			

Leticia Farias Wanderley and Carrie Demmans Epp. Identifying negative language transfer in learner errors using POS information. In Proceedings of the 16th Workshop on Innovative Use of NLP for Building Educational Applications, pages 64–74, Online, April 2021. Association for Computational Linguistics.

Results

Approach	Span	P	R	F1
N-gram baseline	Padded error	0.68	0.32	0.43
	Error + unigram	0.64	0.34	0.45
	Error + bigram			
RNN	Padded error	0.69	0.34	0.46
	Error + unigram	0.67	**0.41**	**0.51**
	Error + bigram			

Leticia Farias Wanderley and Carrie Demmans Epp. Identifying negative language transfer in learner errors using POS information. In Proceedings of the 16th Workshop on Innovative Use of NLP for Building Educational Applications, pages 64–74, Online, April 2021. Association for Computational Linguistics.

Results

Approach	Span	P	R	F1
N-gram baseline	Padded error	0.68	0.32	0.43
	Error + unigram	0.64	0.34	0.45
	Error + bigram	0.66	0.27	0.38
RNN	Padded error	0.69	0.34	0.46
	Error + unigram	0.67	**0.41**	**0.51**
	Error + bigram	**0.70**	0.35	0.46

Leticia Farias Wanderley and Carrie Demmans Epp. Identifying negative language transfer in learner errors using POS information. In Proceedings of the 16th Workshop on Innovative Use of NLP for Building Educational Applications, pages 64–74, Online, April 2021. Association for Computational Linguistics.

Limitation

Part-of-speech tagset

- Using a POS tagset that is common across different languages allowed us to directly compare language structures
- However, this shared tagset is not detailed enough to represent some error types

Limitation: part-of-speech tagset

"It remind me of what I experienced."

Limitation: part-of-speech tagset

"It <u>remind</u> me of what I experienced."

Limitation: part-of-speech tagset

"It **reminds** me of what I experienced."

Limitation: part-of-speech tagset

“It remind me of what I experienced.”

Limitation: part-of-speech tagset

"It remind me of what I experienced."

PRON VERB

Limitation: part-of-speech tagset

“It remind me of what I experienced.”

PRON VERB

PRON: 3rd person singular VERB: non-3rd person singular

Limitation

Part-of-speech representation

- It is not possible to represent all error types with POS tags
- Semantic errors cannot be represented as POS tags sequence

Limitation: part-of-speech representation

“The TV is so important that you can see one in every <u>family</u>.”

Limitation: part-of-speech representation

"The TV is so important that you can see one in every **home**."

Limitation: part-of-speech representation

“The TV is so important that you can see one in every <u>family</u>.”

Limitation: part-of-speech representation

“The TV is so important that you can see one in every family.”

“home” and “family” map to the same Chinese word

Error type analysis

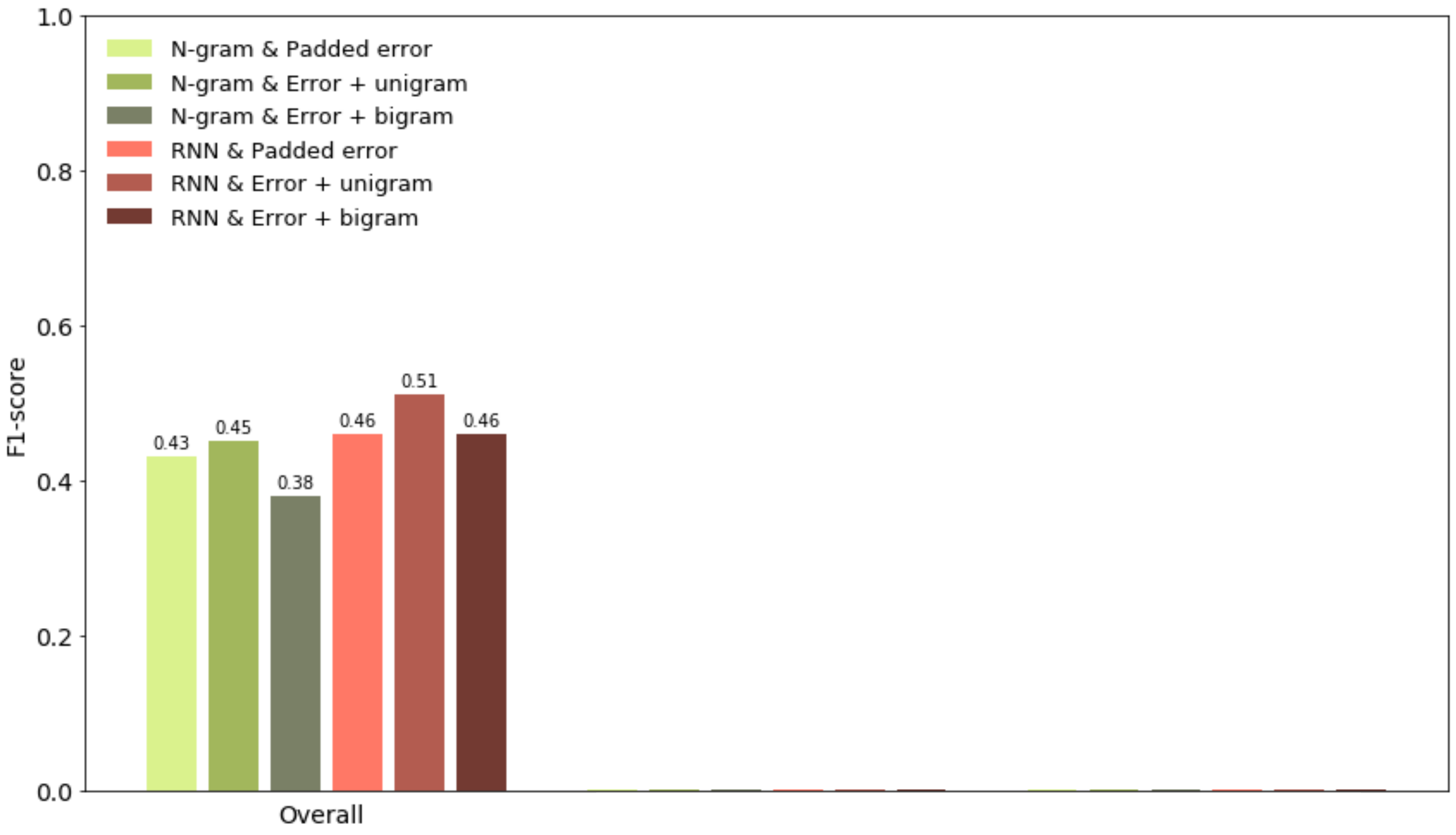
Common errors in the test dataset

- Two of the most common errors made by Chinese native speakers in the FCE dataset are determiner omission and punctuation omission

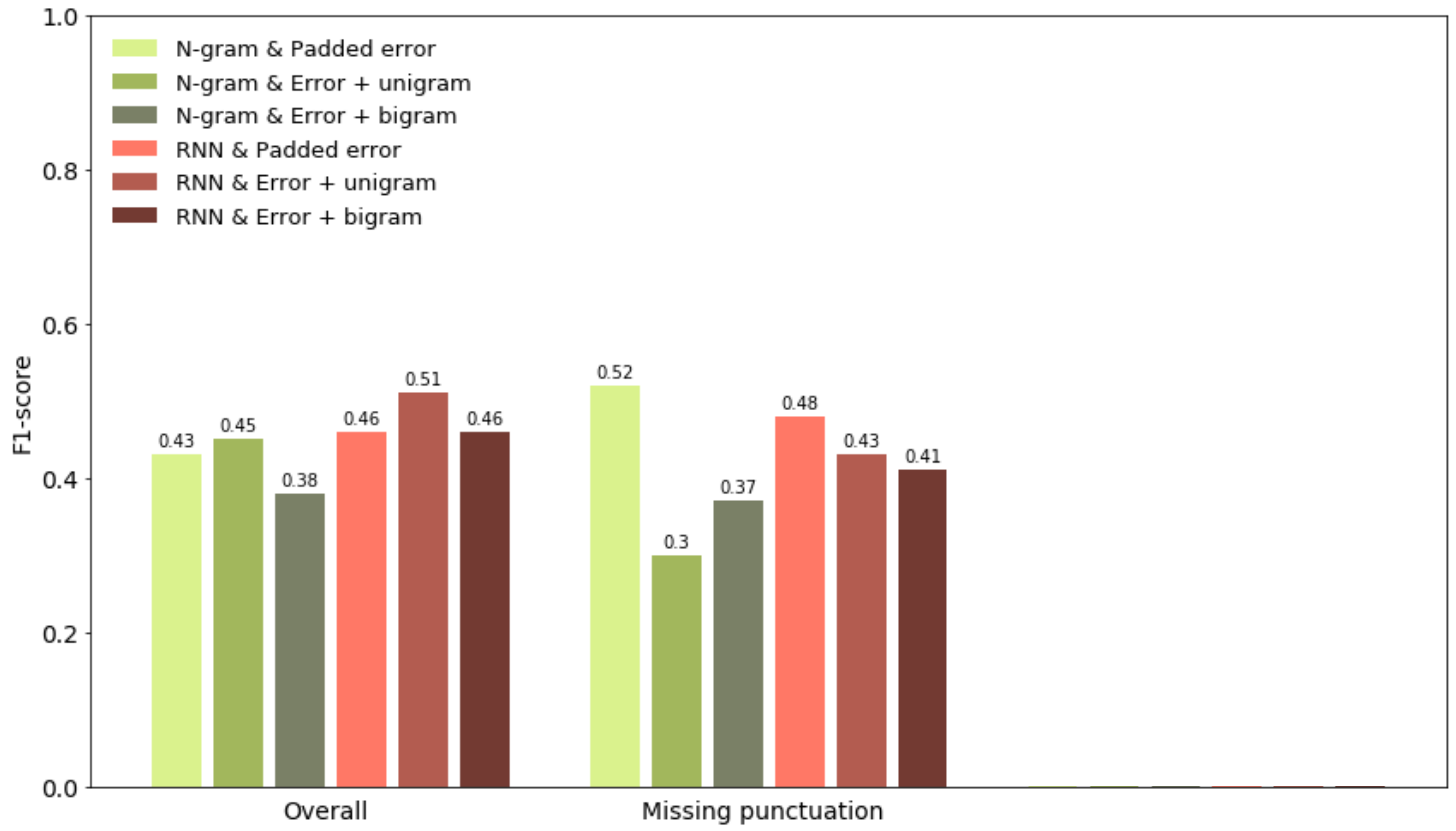
- Both error types are related to learners not using a token when it was necessary

- Both error types are related to negative language transfer

Performance on common errors (F1-score)

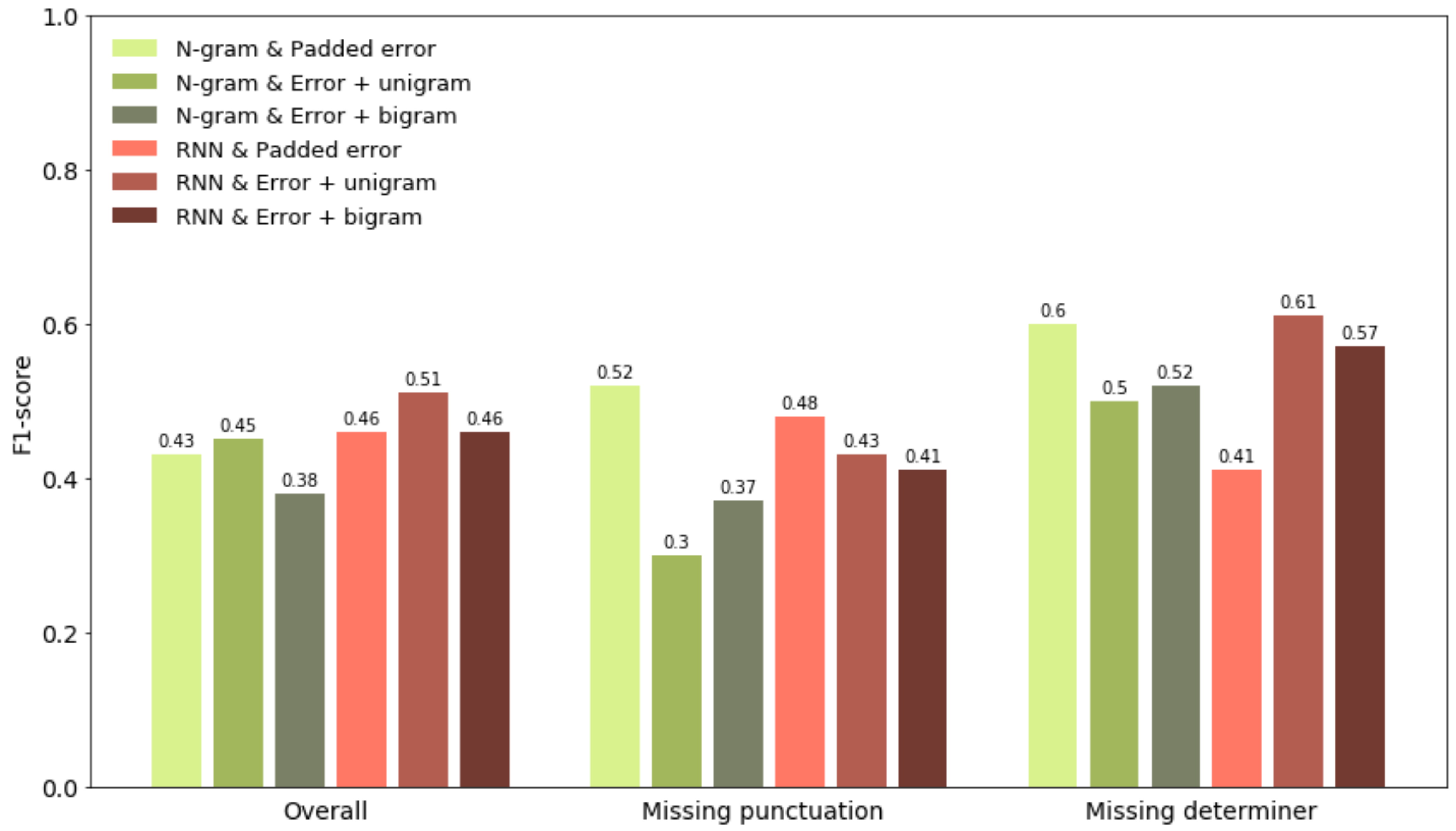


Performance on common errors (F1-score)



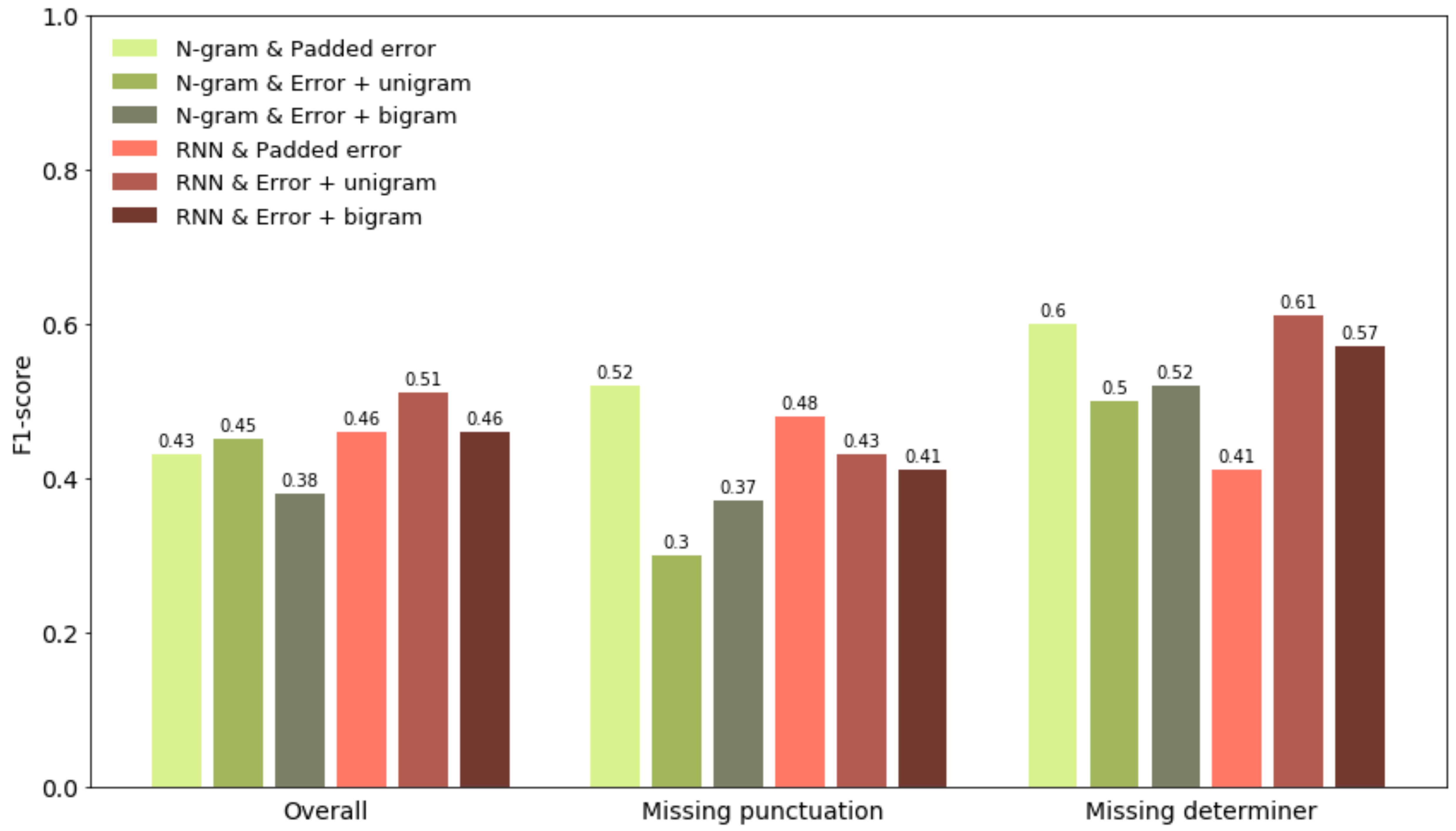
- The padded error span represented missing punctuation errors well in both approaches

Performance on common errors (F1-score)



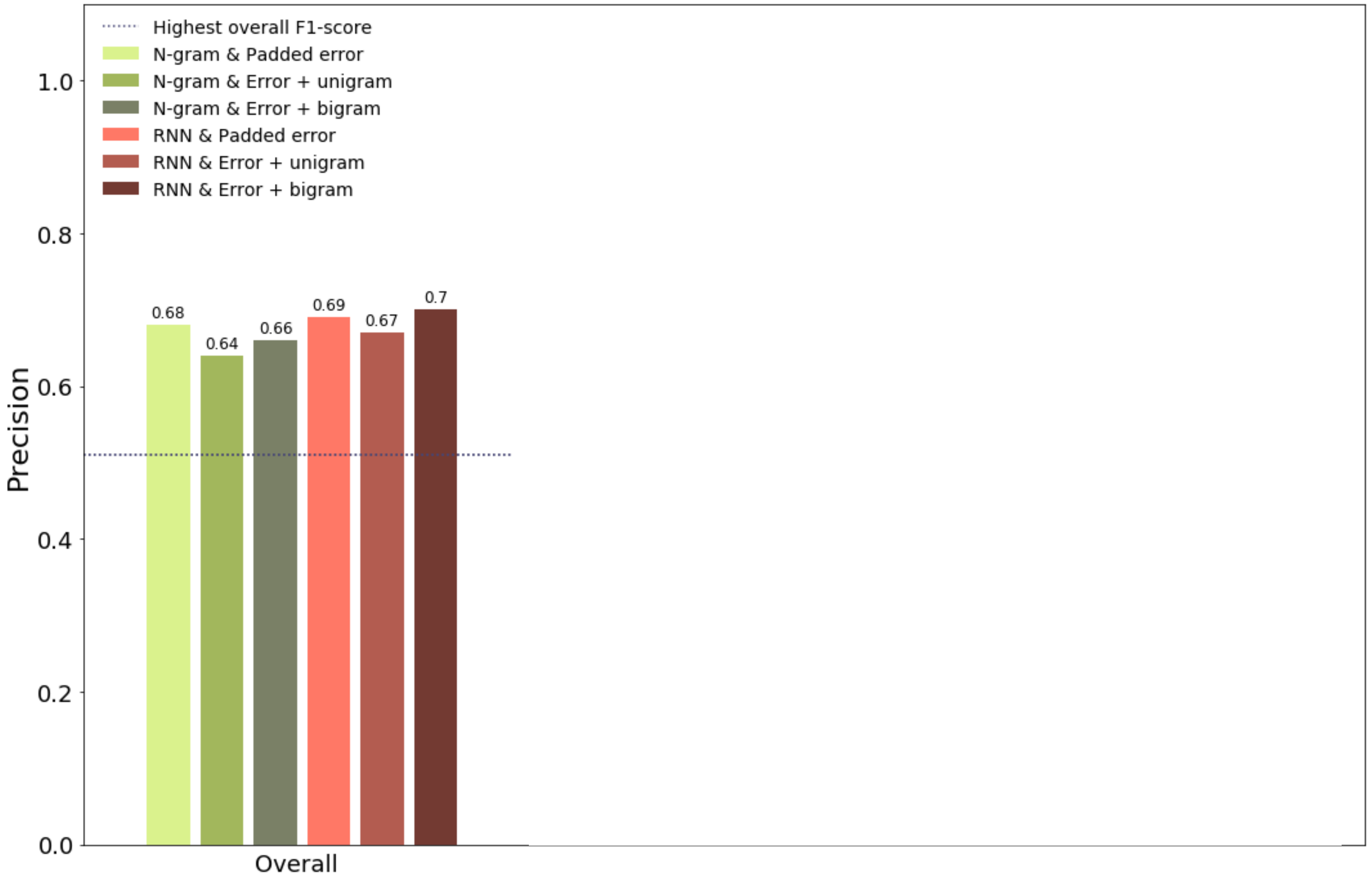
- The padded error span represented missing punctuation errors well in both approaches
- Both error types were better represented by the padded error span in the n-gram approach

Performance on common errors (F1-score)



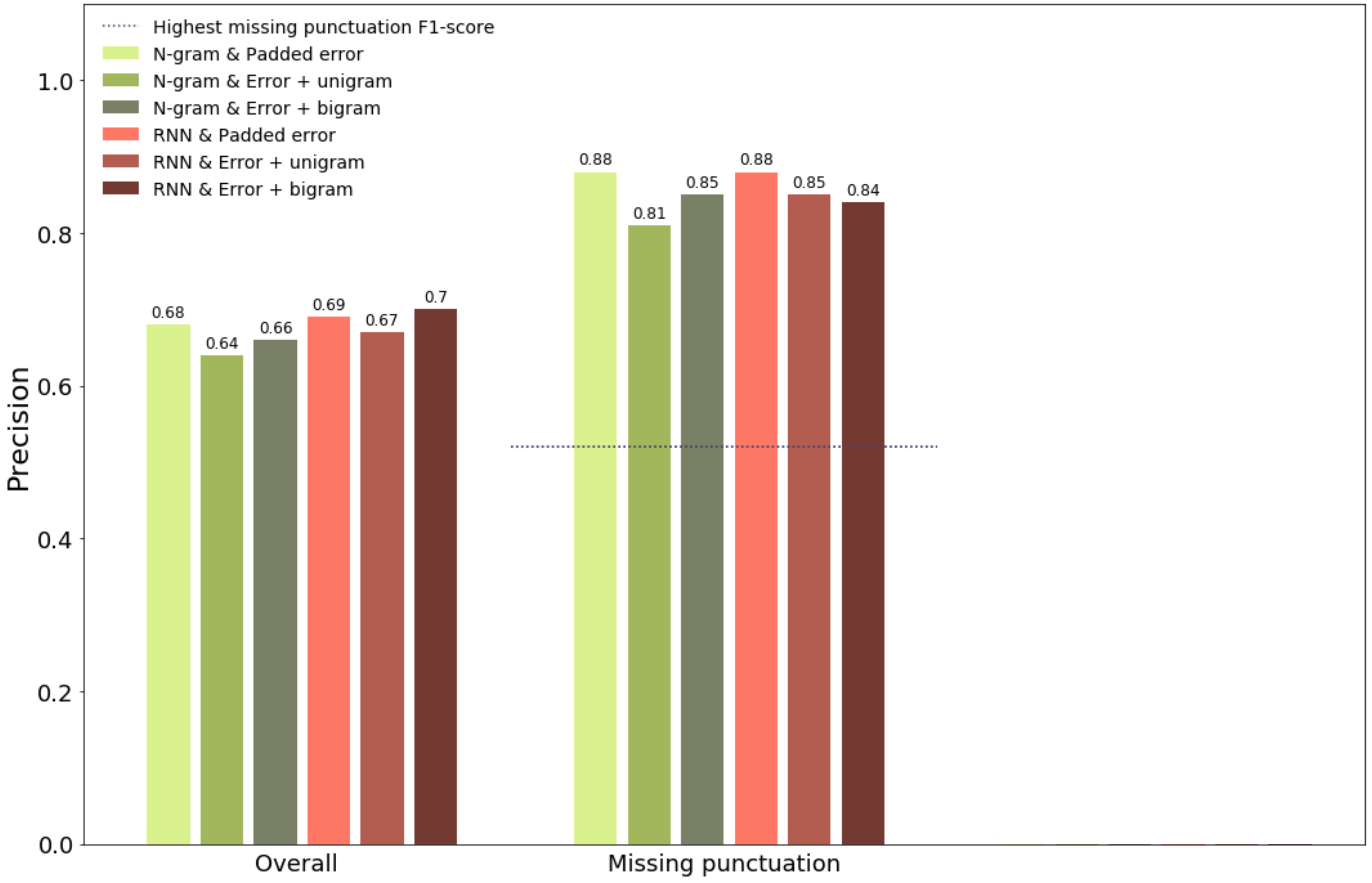
- The padded error span represented missing punctuation errors well in both approaches
- Both error types were better represented by the padded error span in the n-gram approach
- The error + unigram span and RNN approach achieved the highest F1-score among the combinations plotted

Precision on common errors



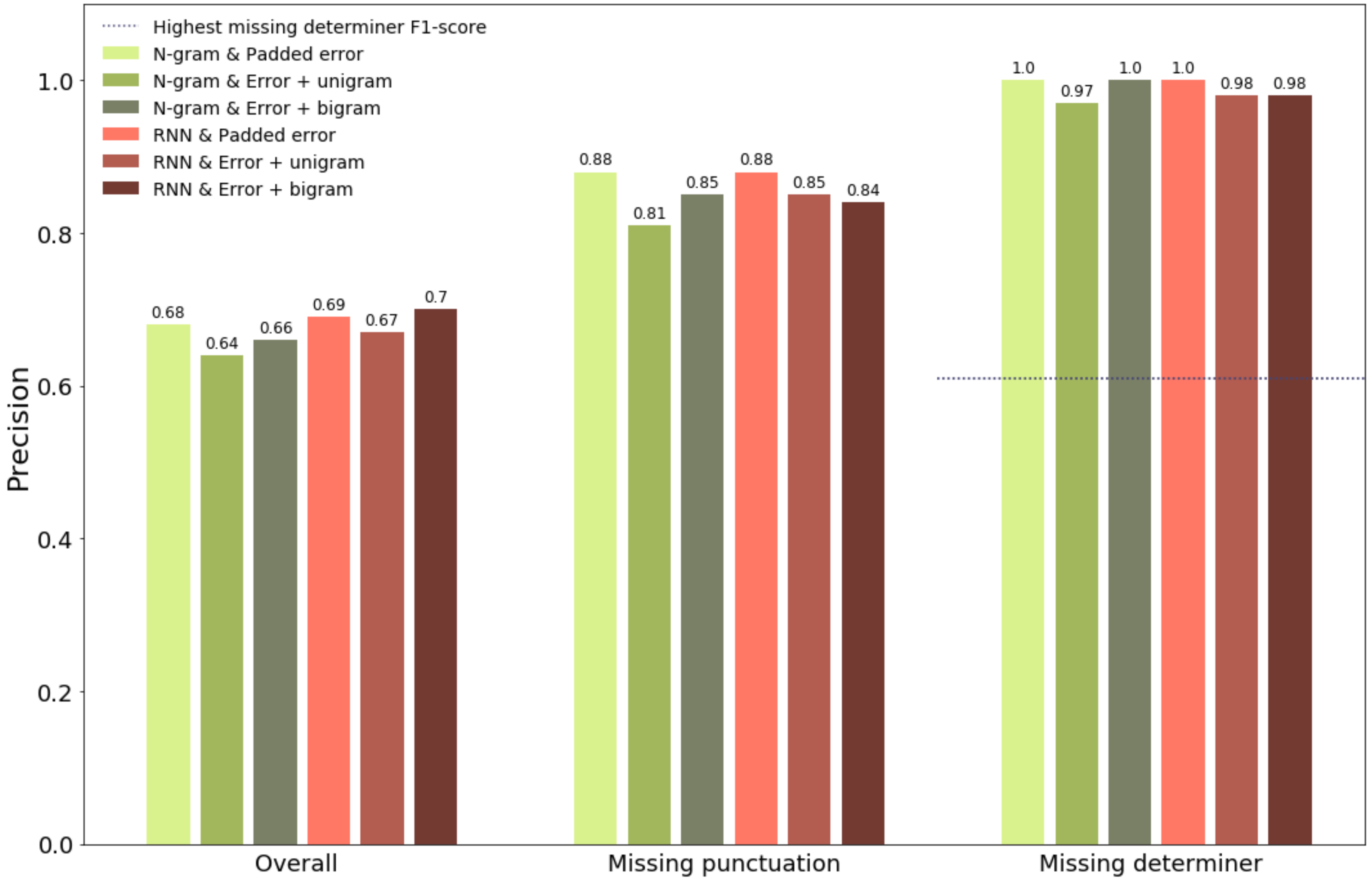
The precision scores achieved by all error span and approach combinations are higher than their equivalent F1-scores

Precision on common errors



The precision scores achieved by all error span and approach combinations are higher than their equivalent F1-scores

Precision on common errors



The precision scores achieved by all error span and approach combinations are higher than their equivalent F1-scores

Future steps

- Integrate models into a writing assistant and provide error feedback enhanced with negative language transfer information

- Design and conduct a user study to understand the impact (if any) of negative language transfer feedback for language learners

Leticia Farias Wanderley and Carrie Demmans Epp. Identifying negative language transfer in writing to increase English as a Second Language learners' metalinguistic awareness. In Companion Proceedings 10th International Conference on Learning Analytics & Knowledge (LAK20), pages 722–725, March 2020.

Implication: language standardisation

- Both training and test data were annotated according to a set of formal English grammar rules
- By highlighting errors that do not follow the structures found on the training data, we may be imposing a specific writing style to language learners

Implications

- Both training and test data were annotated according to a set of formal English grammar rules
- By highlighting errors that do not follow the structures found on the training data, we may be imposing a specific writing style to language learners
- This may lead to a standardisation of English teaching and learning as it doesn't allow for other English varieties

Implications

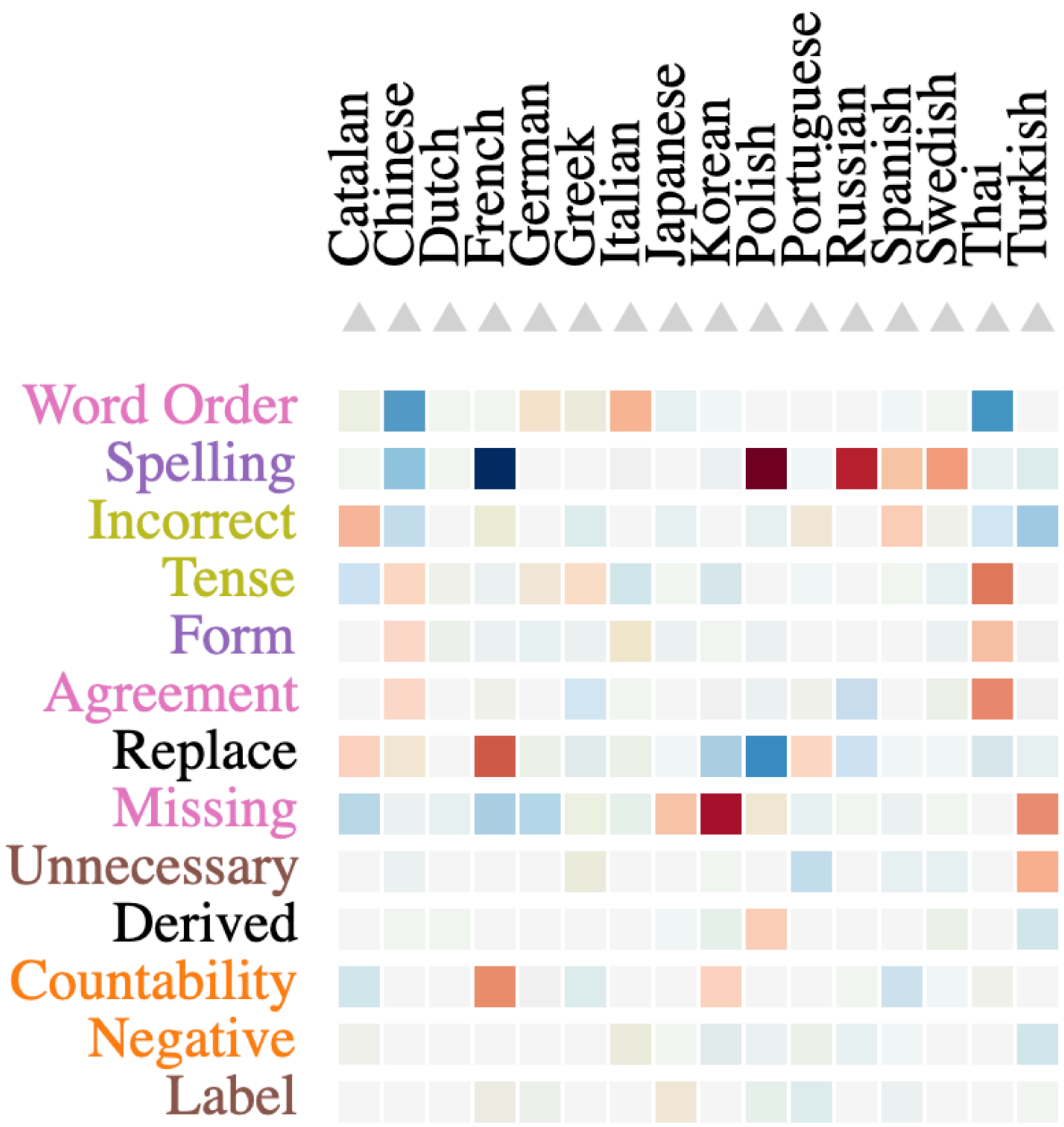
- Both training and test data were annotated according to a set of formal English grammar rules
- By highlighting errors that do not follow the structures found on the training data, we may be imposing a specific writing style to language learners
- This may lead to a standardisation of English teaching and learning as it doesn’t allow for other English varieties
- However, explaining why learners are making certain mistakes is still valuable as it can help them better understand the language they’re learning

Summary

- Introduced the task of negative language transfer identification in learner errors
- Approached the task with a method that uses syntactic information from parallel textual data to identify structural negative language transfer
- Built a negative language transfer dataset with errors made by Chinese native speakers

Extra slides

Error type distribution across languages



Mariana Shimabukuro, Jessica Zipf, Mennatallah El-Assady, and Christopher Collins. H-matrix: Hierarchical matrix for visual analysis of cross-linguistic features in large learner corpora. In Proceedings of the IEEE Conference on Information Visualization (short papers), 2019.

Negative language transfer classification

- Baseline model uses error types to predict the negative language transfer label
- Investigate if syntactic features extracted from the errors improve classification performance
  - Part-of-speech tags
  - Error length

Leticia Farias Wanderley, Nicole Zhao, and Carrie Demmans Epp. Negative language transfer in learner English: A new dataset. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 3129–3142, Online, June 2021. Association for Computational Linguistics.

Negative language transfer classification

Features	Acc	P	R
Error types	0.72	0.79	0.73
Error types + syntactic features	**0.78**	**0.82**	**0.79**

Leticia Farias Wanderley, Nicole Zhao, and Carrie Demmans Epp. Negative language transfer in learner English: A new dataset. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pages 3129–3142, Online, June 2021. Association for Computational Linguistics.