CoNLL 2019

Multilingual model using cross-task embedding projection

Jin Sakuma
The University of Tokyo

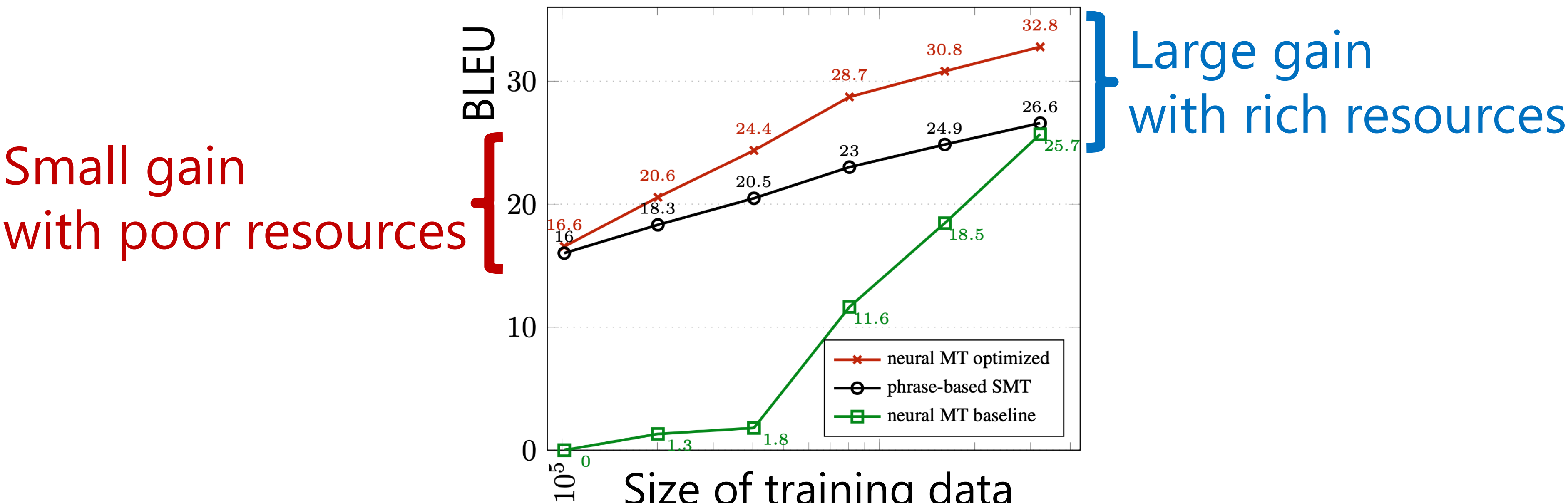
Naoki Yoshinaga
Institute of Industrial Science, the University of Tokyo

Needs for resources in deep learning

Improvements in various tasks by deep learning

- **Task-specific representation learning** needs more data

Translation performance (DE->EN) [Sennrich+ 2019]



Need massive data for **every pair of task and lang**

Gaps in available resources across langs

Among 7097 languages in the world [Simons+ 2018], massive resources are obtainable only in few

- Universal dependencies project covers only 76 [Nivre+ 201X]

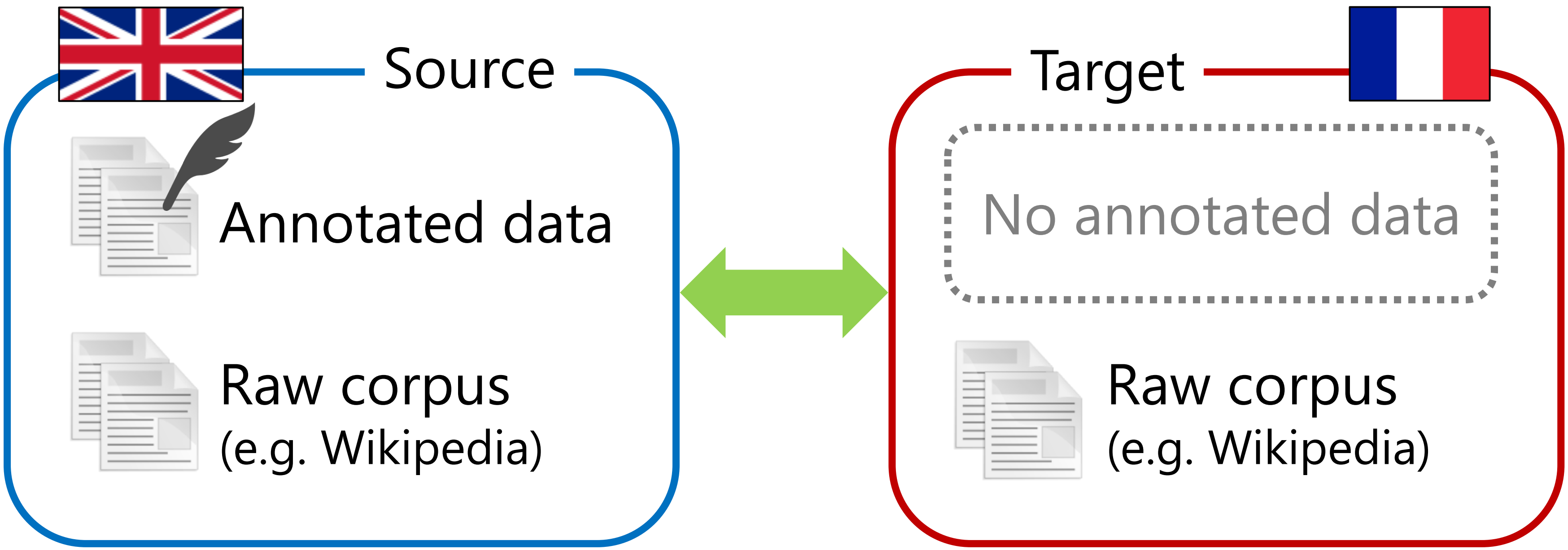
Large gap in model performances among languages

Can we exploit resources of resource-rich languages for training in resource-poor languages?

Problem settings of this study

Available resources:

- Labeled data for training in the source language
- Raw corpora in both languages



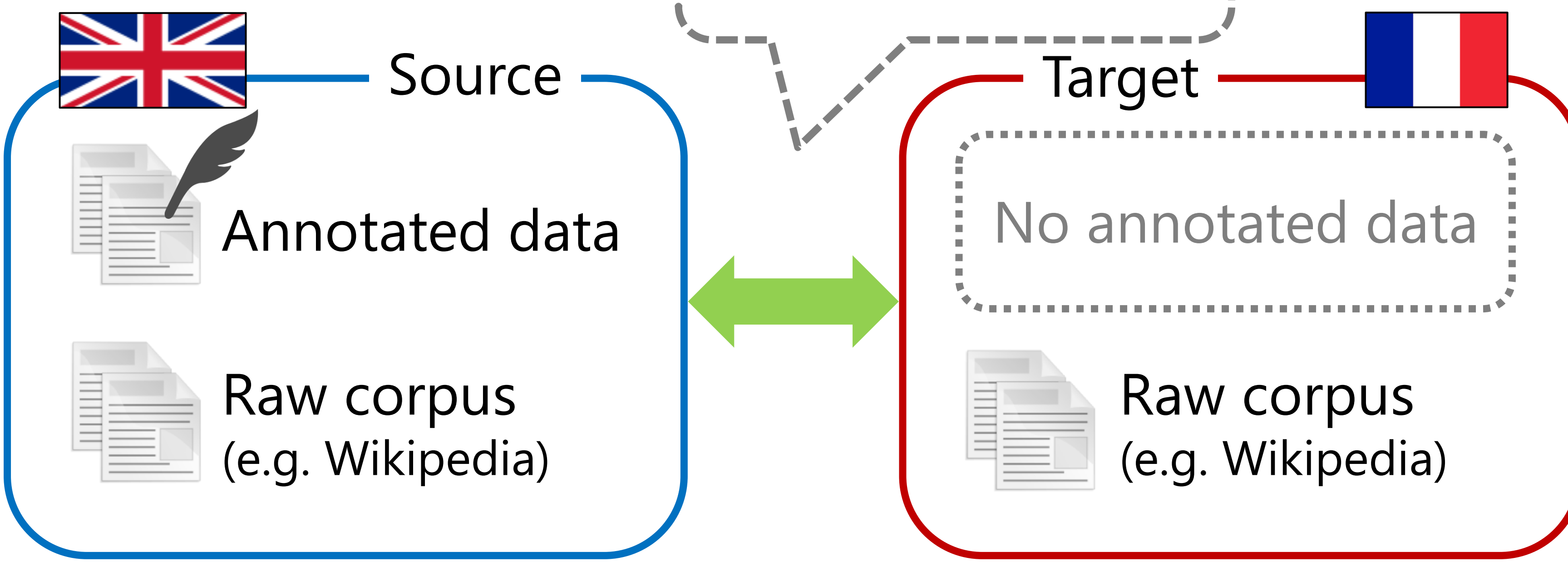
Applicable to various target languages and tasks

Problem settings of this study

Available resources:

- Labeled data for trainin ... ge
- Raw corpora in both lan

No cross-lingual resources

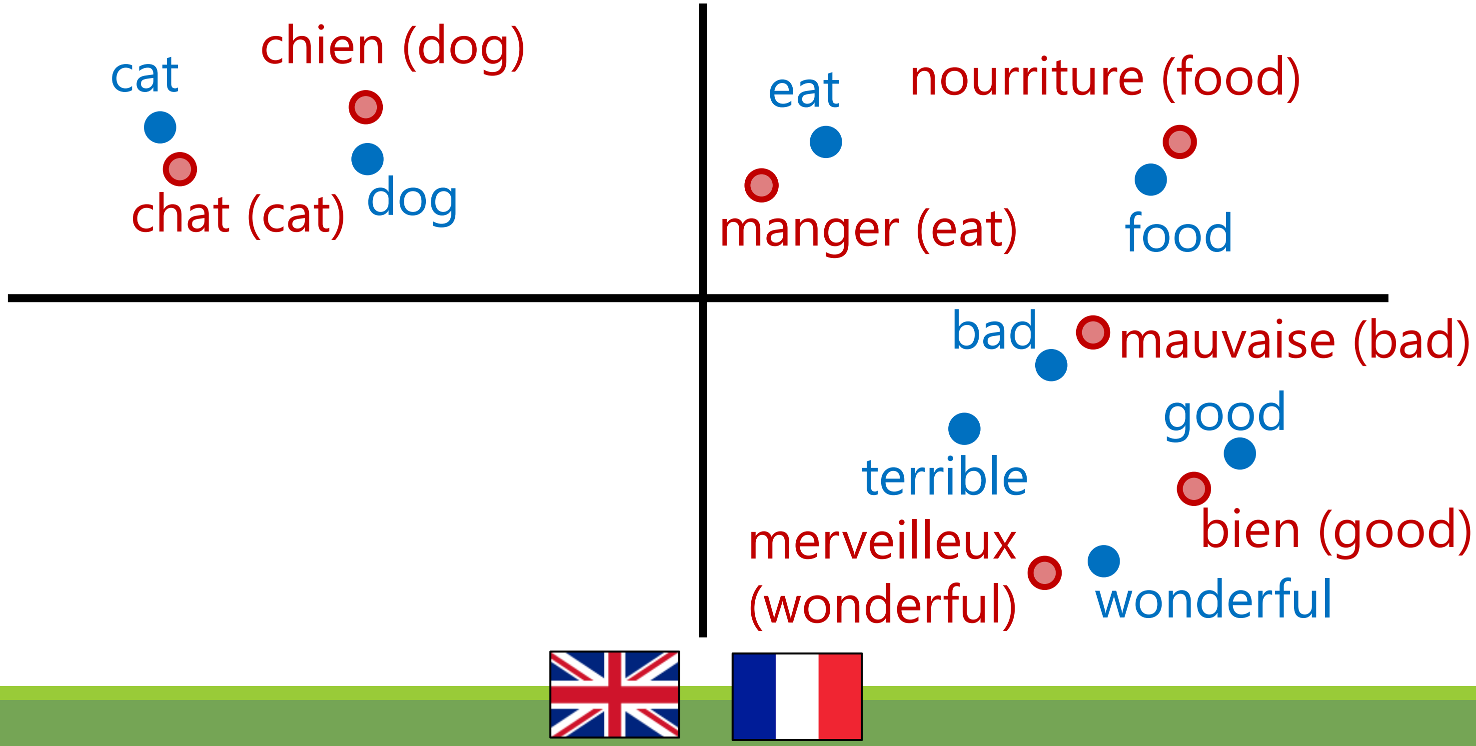


Applicable to various target languages

Preliminary:
Cross-lingual word embeddings (CLWE)

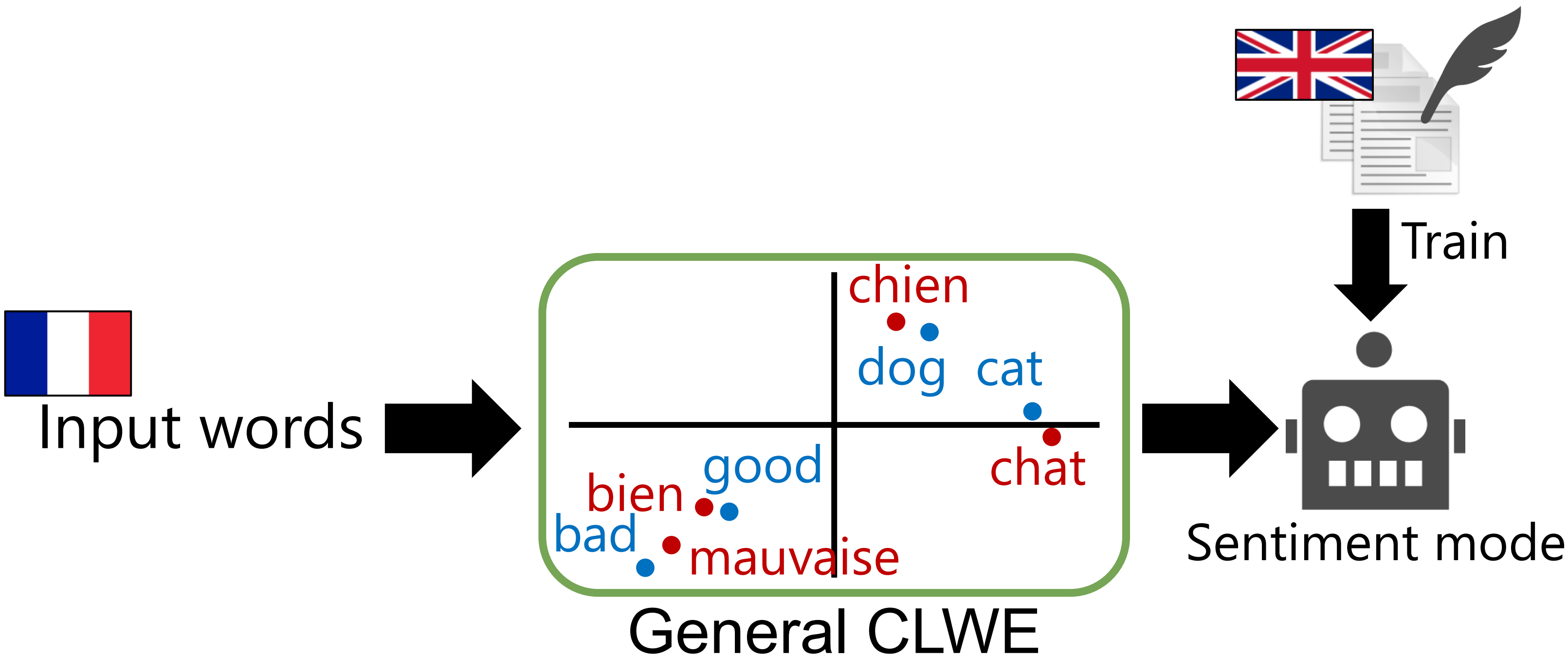
Language-independent representation of words
[Mikolov+ 13]

- Words from two lang. are represented in a shared space
- Similar words from different languages are close



Existing multilingual models

Fix the emb. layer to general CLWE during training
[Duong+ 17, Chen+ 18]

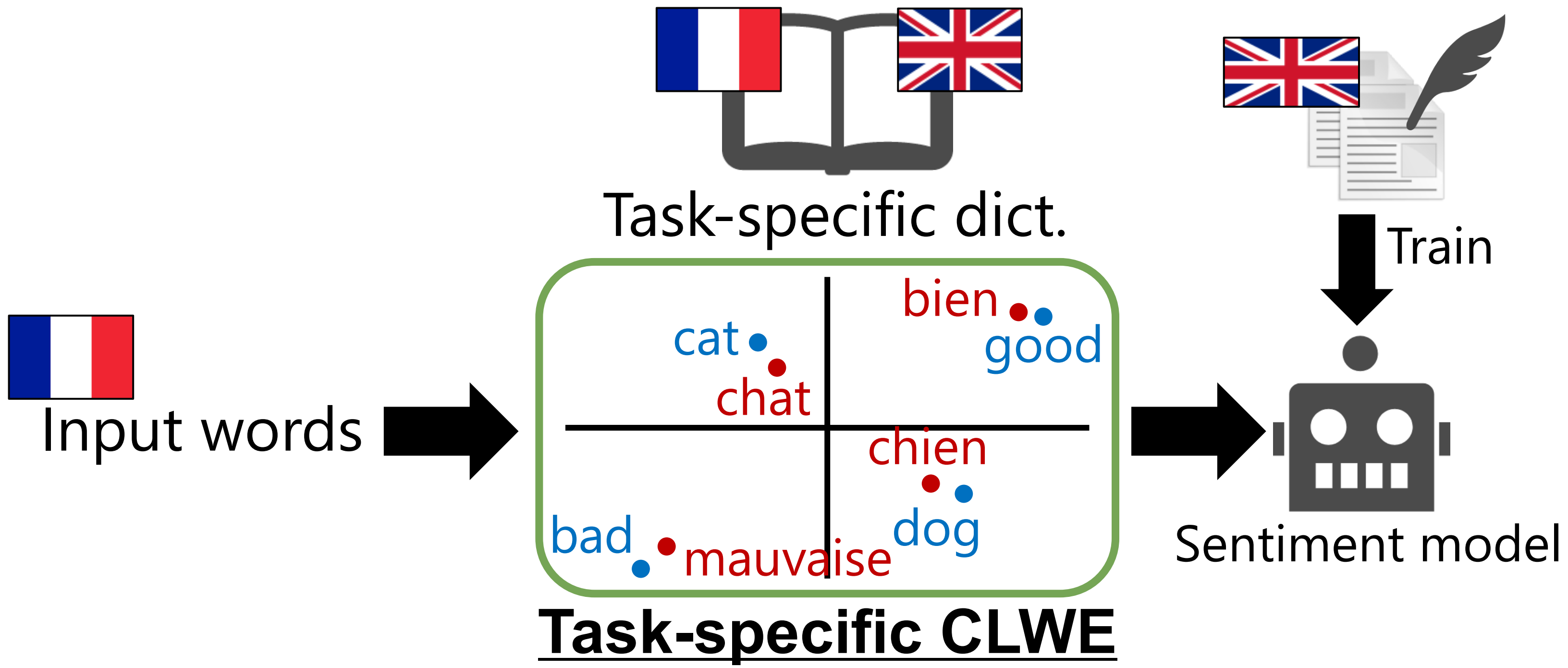


😃 Enables cross-lingual transfer

😰 The embedding layer is not optimized for the task

Related work:
Task-specific CLWE with specialized dict.

Utilize task-specific bilingual dict. to obtain CLWE
[Gouws+ 15]



😃 The embedding layer is optimized for the task

😰 Additional cross-lingual resources are required

Related work:
Task-specific CLWE with specialized dict.

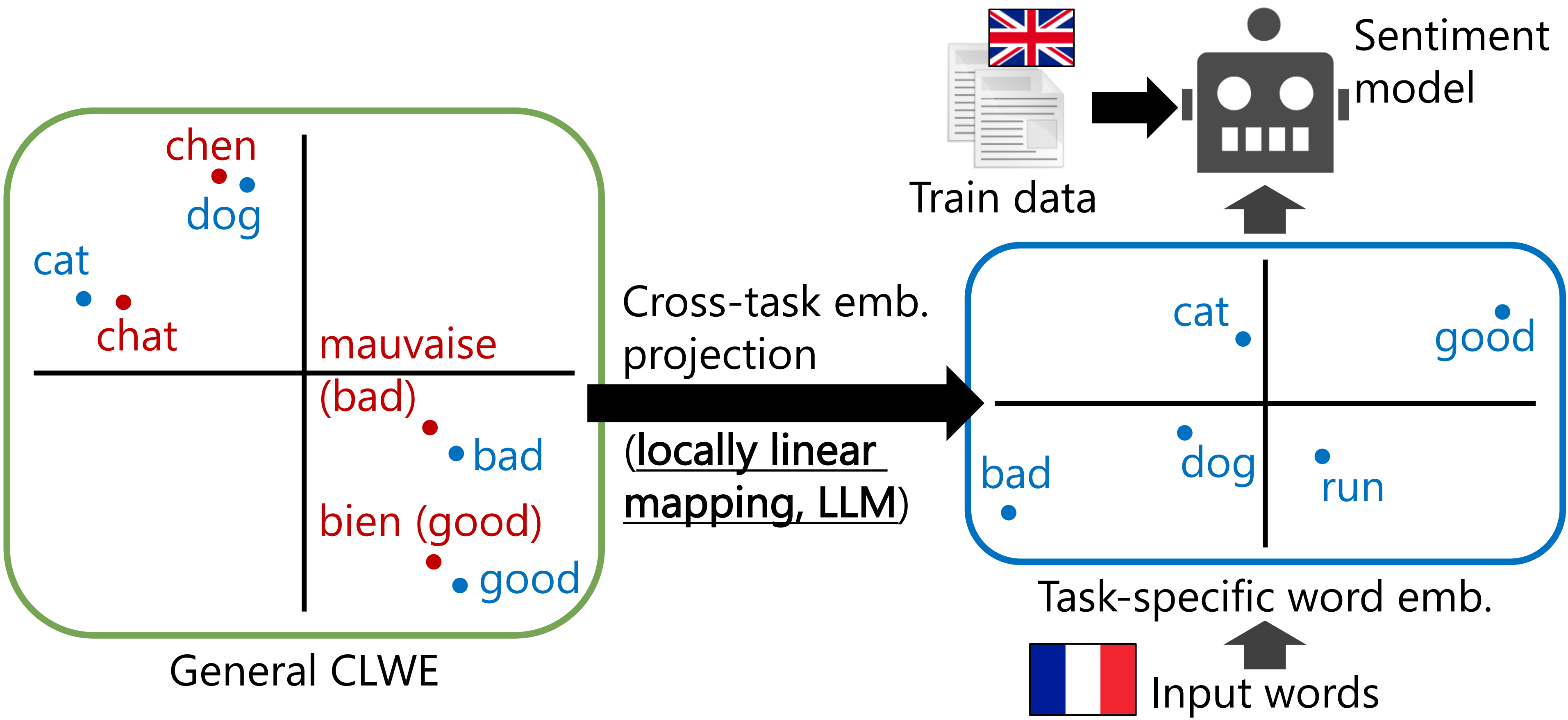
Utilize task-specific bilingual dict. to obtain CLWE [Gouws+ 15]

In this study:
Obtain task-specific CLWE without relying on any cross-lingual resources

- The embedding layer is optimized for the task
- Requires additional cross-lingual resources

Proposal:
Multilingual model with task-spec emb.

Project general CLWE to the emb. layer optimized for the task by **cross-task embedding projection**

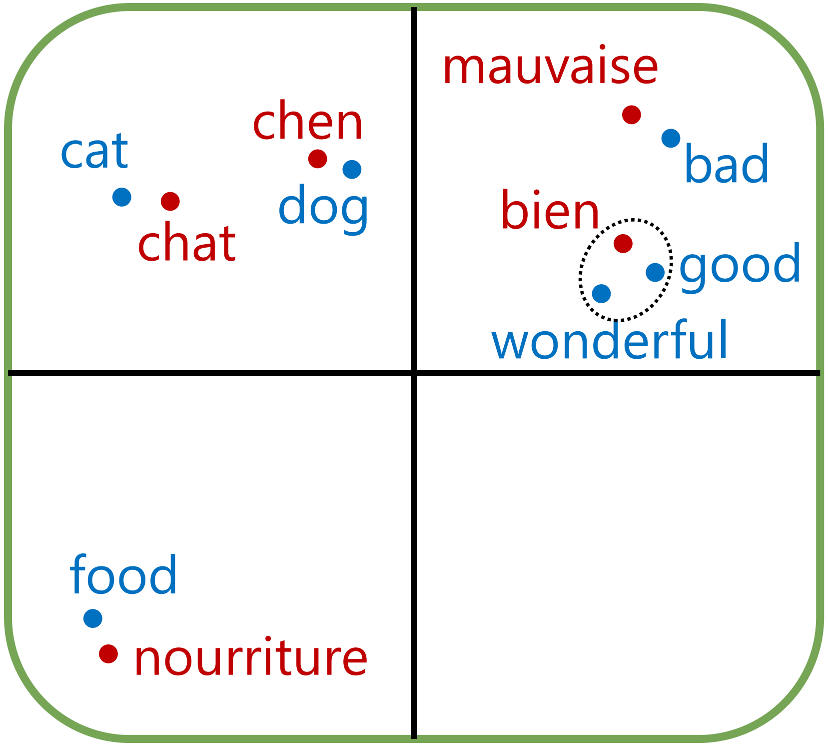


Locally linear mapping:

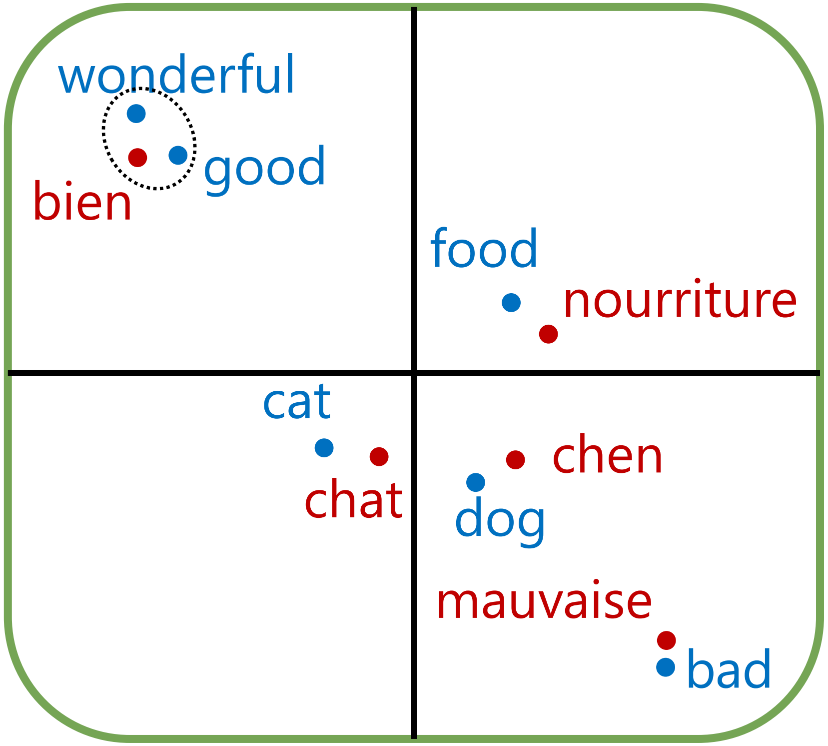
Idea: local topology of embeddings

Assumption:

- Words <u>**adequately close**</u> in the general CLWE are also close in task-specific space



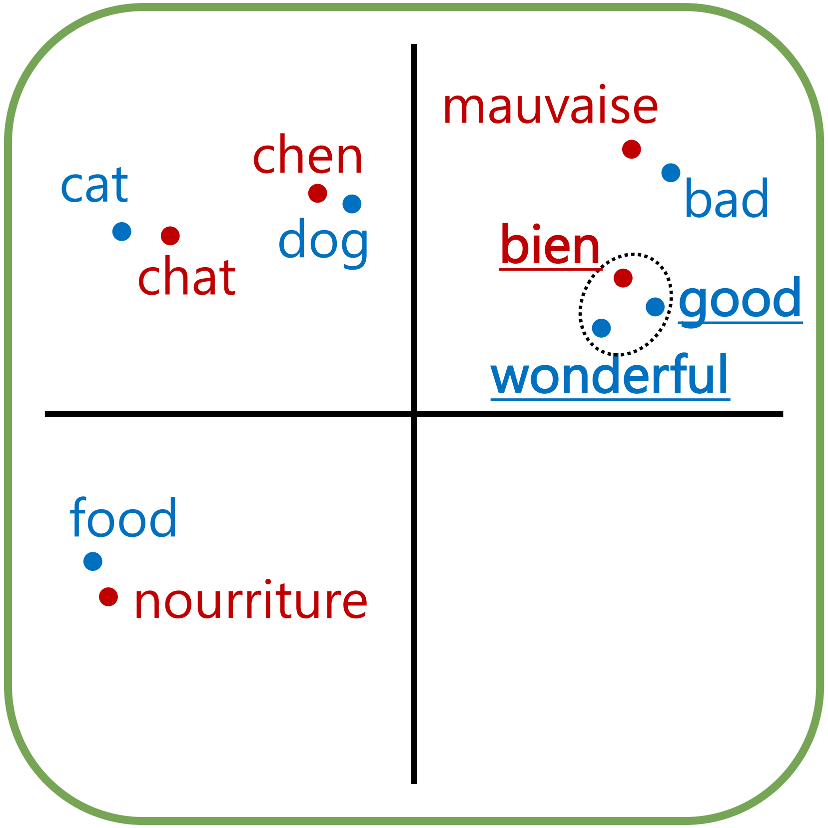
General CLWE



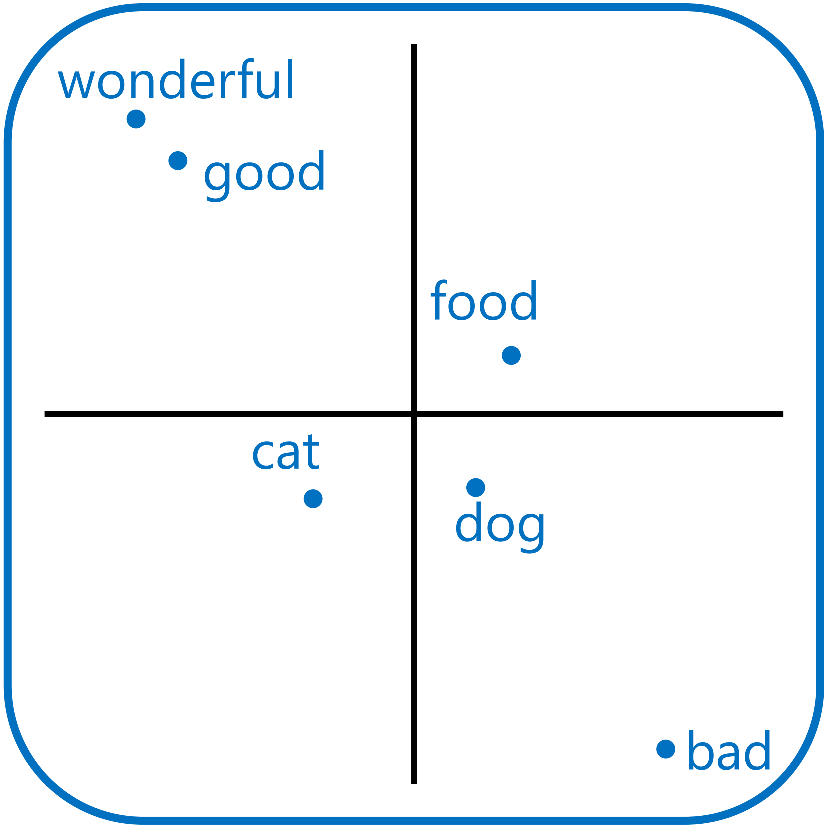
(Ideal) task-specific CLWE

Locally linear mapping:
Step 1: selecting nearest neighbors

For each **target word (bien)**, select ***k*-nearest neighbors** in the general CLWE



General CLWE

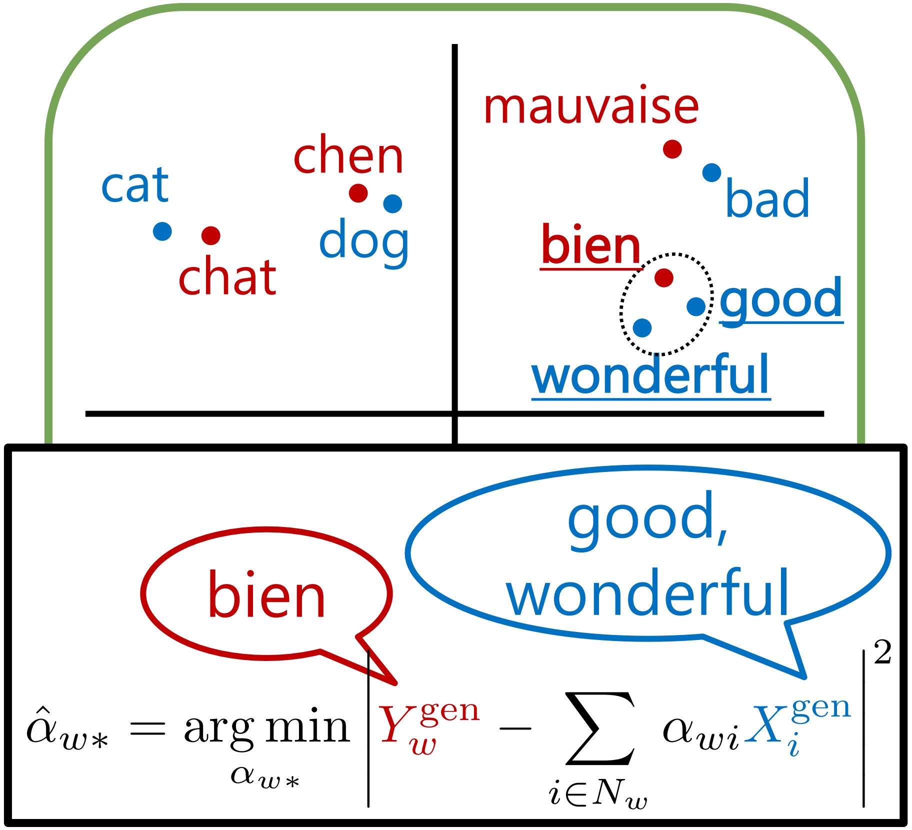


Task-specific word emb.

Locally linear mapping:

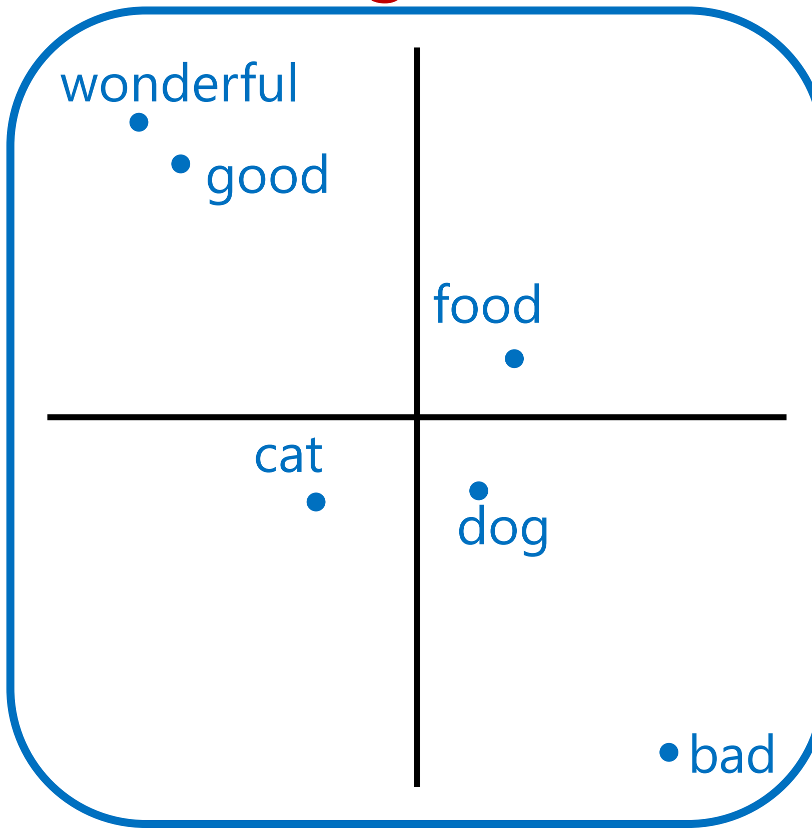
Step 2: local topology in general space

In general CLWE, learn linear combination of **nearest neighbors** that reconstructs the **the target word**



$$\hat{\alpha}_{w*} = \arg\min_{\alpha_{w*}} \left| Y_w^{\text{gen}} - \sum_{i \in N_w} \alpha_{wi} X_i^{\text{gen}} \right|^2$$

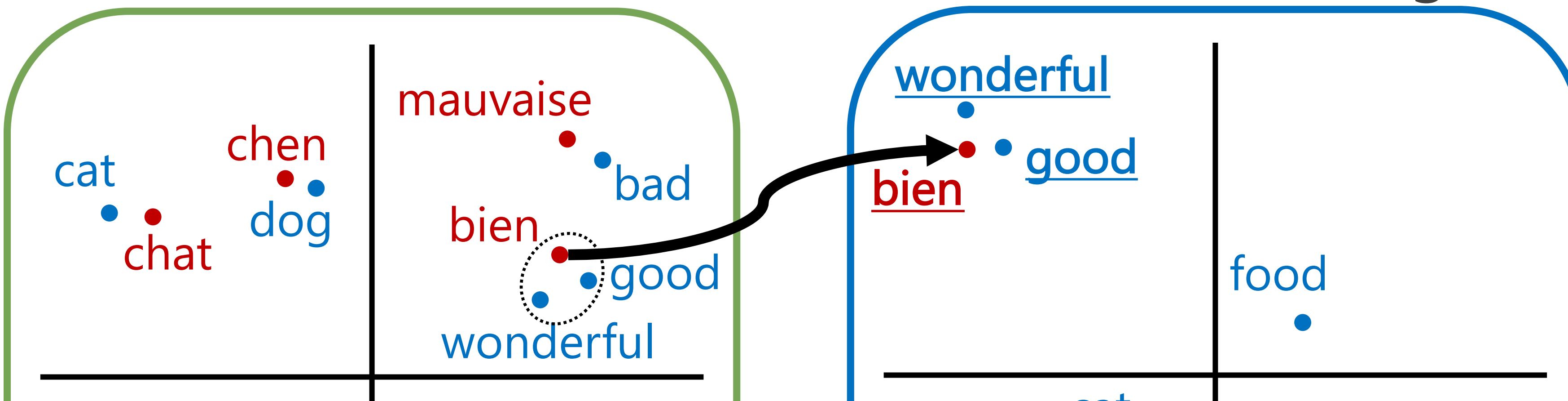
General CLWE



Task-specific word emb.

Locally linear mapping:
Step 3: task-specific word embeddings

Compute **task-specific word emb. of the target word** as the linear combination with the induced weights



bien

good, wonderful

$$\hat{\alpha}_{w*} = \underset{\alpha_{w*}}{\arg\min} \left| Y_w^{\text{gen}} - \sum_{i \in N_w} \alpha_{wi} X_i^{\text{gen}} \right|^2$$

General CLWE

bien

good, wonderful

$$Y_w^{\text{spec}} = \sum_{i \in N_w} \hat{\alpha}_{wi} X_w^{\text{spec}}$$

Task-specific word emb.

Proposal:
Hyperparameter search

Dev. set in the target language is required to tune the hyperparameter k (size of nearest neighbors)

Tuning to the task (no additional resources)

Assume the best k is **independent of language**

Apply LLM to the embeddings of the source language and evaluate on the dev. set of the source language

(Tuning to the task/language)

Utilize small development set (100 examples) of the target language

Experimental setup (1/2)

Goal:

- Does out task-specific word embeddings improve the multilingual model?

Task:

- Topic classification task (and sent. analysis)

Languages:

- Source language: English (en)
- Target languages:
  - Danish (da), Italian (it), French (fr), Swedish (sv)

Datasets:

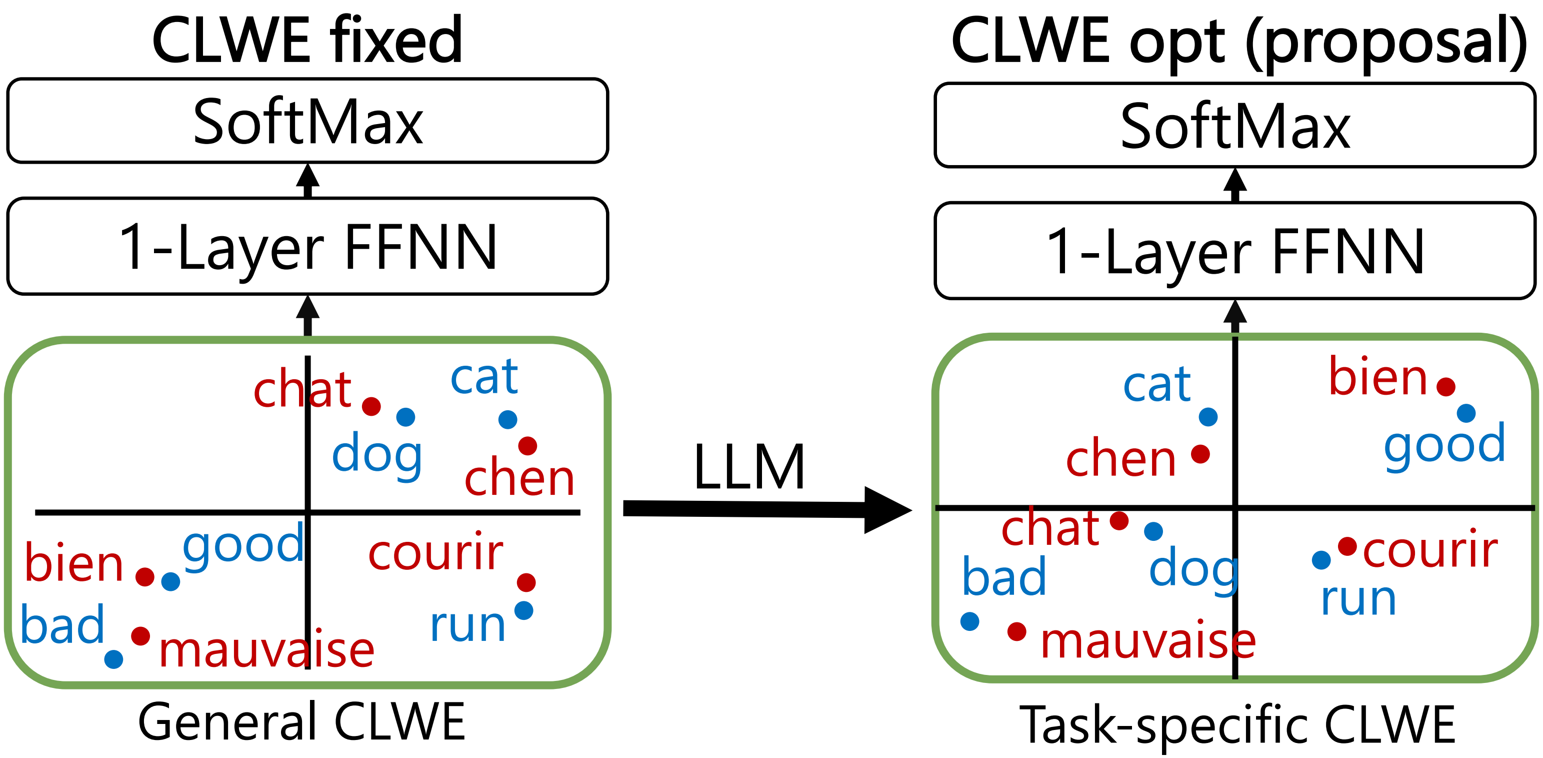
- RCV1/2 dataset (four topics)

Experimental setup (2/2):

Models to compare

Compare the following two models to evaluate the effect of task-specific CLWE

- Experiments on more models on the paper



Results:

Topic classification task

Classification accuracies in four languages

Method	k-tuning	en-da	en-it	en-fr	en-sv
CLWE fixed	-	0.621	0.535	0.772	0.816
CLWE opt (Proposed)	task	0.672	**<u>0.623</u>**	**<u>0.885</u>**	**<u>0.831</u>**
CLWE opt (Proposed)	task/lang	**<u>0.687</u>**	0.615	0.879	0.830

- CLWE opt outperforms the baseline
- Tuning k for task and language is not necessary

Conclusion and future work

Conclusion

- Proposed a method to build a multilingual model with <u>task-specific word embeddings</u>
- Evaluated our method on real tasks and <u>confirmed its effectiveness</u>

Future work

- Evaluate this method on wider range of tasks, languages, and models
- Further improve the quality of locally linear mapping