Cognition-aware Cognate Detection

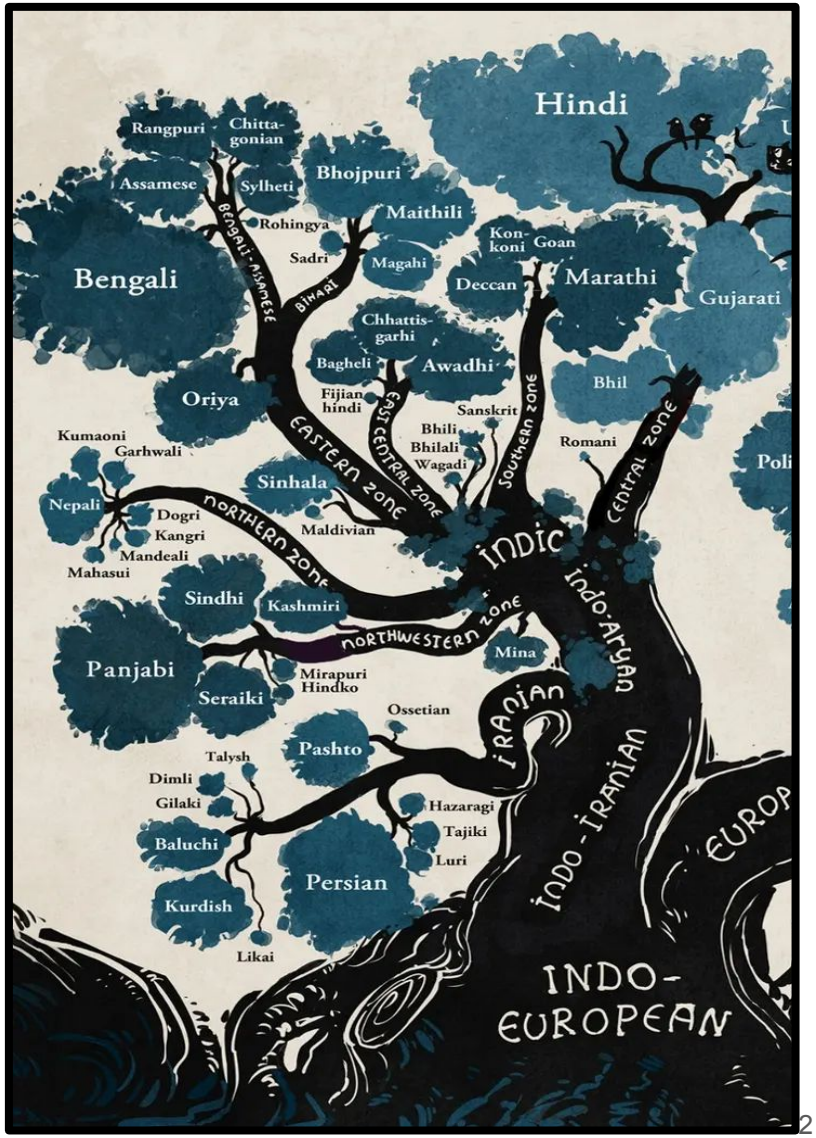
Diptesh Kanojia | Prashant K. Sharma | Sayali Ghodekar |
Pushpak Bhattacharyya | Gholamreza Haffari | Malhar Kulkarni

European Chapter of the Association for Computational Linguistics

An Indian-Australian research partnership

Cognate Detection : Motivation

- Cognates represent a large chunk of the shared vocabulary among language pairs.
- We conduct this experiment for an Indian language pair Hindi - Marathi, which is a known closely related pair.
- Previously, the task of Cognate Detection has shown to help the downstream tasks of Machine Translation via word alignment (Kondrak, 2005)
- Cognitive Psycholinguistic based features have also shown to improve various NLP tasks (Mishra et. al., 2016)



Cognition Aware Cognate Detection [1 /2]

Problem Statement

Key Question: Do cognitive (gaze) features help in cognate detection ?

GOALS

- **Collect gaze behaviour data** for the task of identifying cognates vs. non-cognates for a sample set.
- **Extract gaze features** from the collected gaze data.
- **Predict gaze features** for the unseen samples.
- Perform the **task of cognate detection** over both sets.

INPUT

Cognate Challenge Dataset (Kanojia et. al., 2020)
+
Traditional features
+
Gaze data

OUTPUT

Cognates (1) / Non-Cognates (0)

Cognition Aware Cognate Detection [2 / 2]

- **Vector Representation:**
  - W1,W2, D1, D2, E1, E2
  - From *Cognate Challenge Dataset (Kanojia et. al., 2020)*
- **Traditional features**
  - Phonetic, Lexical etc.
- **Gaze Features**
  - g1, g2, g3,.....g_n
  - from collected data

INPUT

Vector Representation
+
Traditional features
+
Gaze data

OUTPUT

Cognates (1) /
Non-Cognates (0)

Sub-Problem: Predicting Cognitive Features

Problem Statement

GOAL

- Using the collected gaze data, predict gaze features for the unseen samples of cognates and non-cognates.

- **Vector Representation:**
  - W1,W2, D1, D2, E1, E2
- **Traditional features**
  - Phonetic, Lexical etc.
- **Gaze Features)**
  - g1, g2, g3,.....g_n
  - from collected data

INPUT

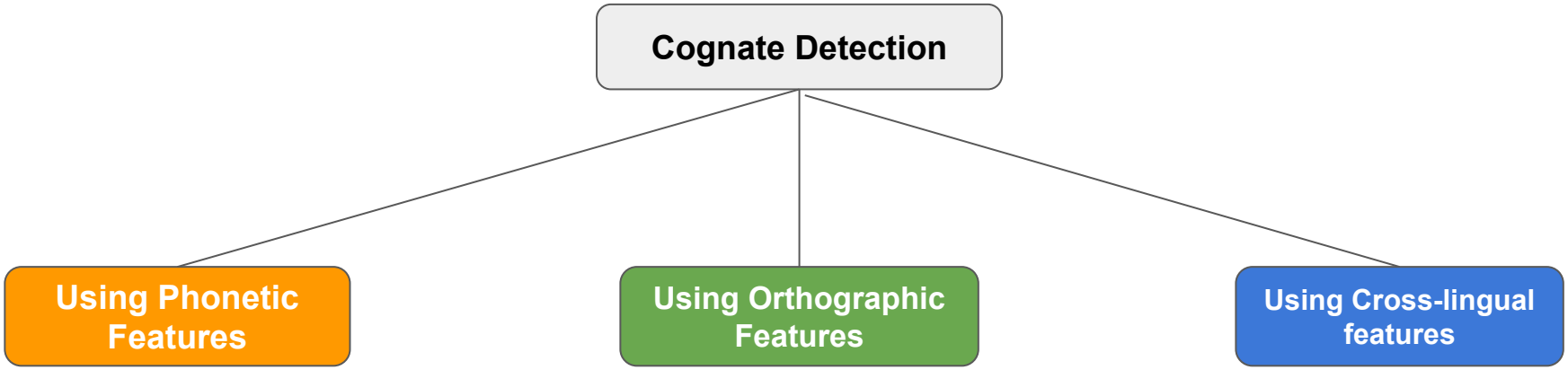
Vector Representation
+
Traditional features
+
Gaze Features
(from collected data)

OUTPUT

Gaze Features
(on unseen data)

G1, G2, G3,.....G_n

Literature Survey



Rama, T. (2016, December). Siamese convolutional networks for cognate identification. *(COLING 2016)*

Kondrak, G. (2001). Identifying cognates by phonetic and semantic similarity. (NAACL 2001)

Ciobanu, A. M., & Dinu, L. P. (2014, June). Automatic detection of cognates using orthographic alignment. (ACL 2014)

Mulloni, Andrea, and Viktor Pekar. "Automatic Detection of Orthographics Cues for Cognate Recognition." (LREC 2006)

Kanojia et. al. Utilizing cross-lingual word embeddings for Cognate Detection (COLING 2020)

Merlo, P., & Rodriguez, M. A. (2019, November). Cross-lingual word embeddings and the structure of the human bilingual lexicon. (CoNLL 2019)

Dataset Collection Setup

Annotator Info

- Nine annotators
- Native Marathi speakers
  (who understand Hindi)
- Education Level
  - At least College Graduates
- Experiments conducted with a host always at the side
- SR Research EyeLink 1000
  (used at 500 Hz sampling rate)

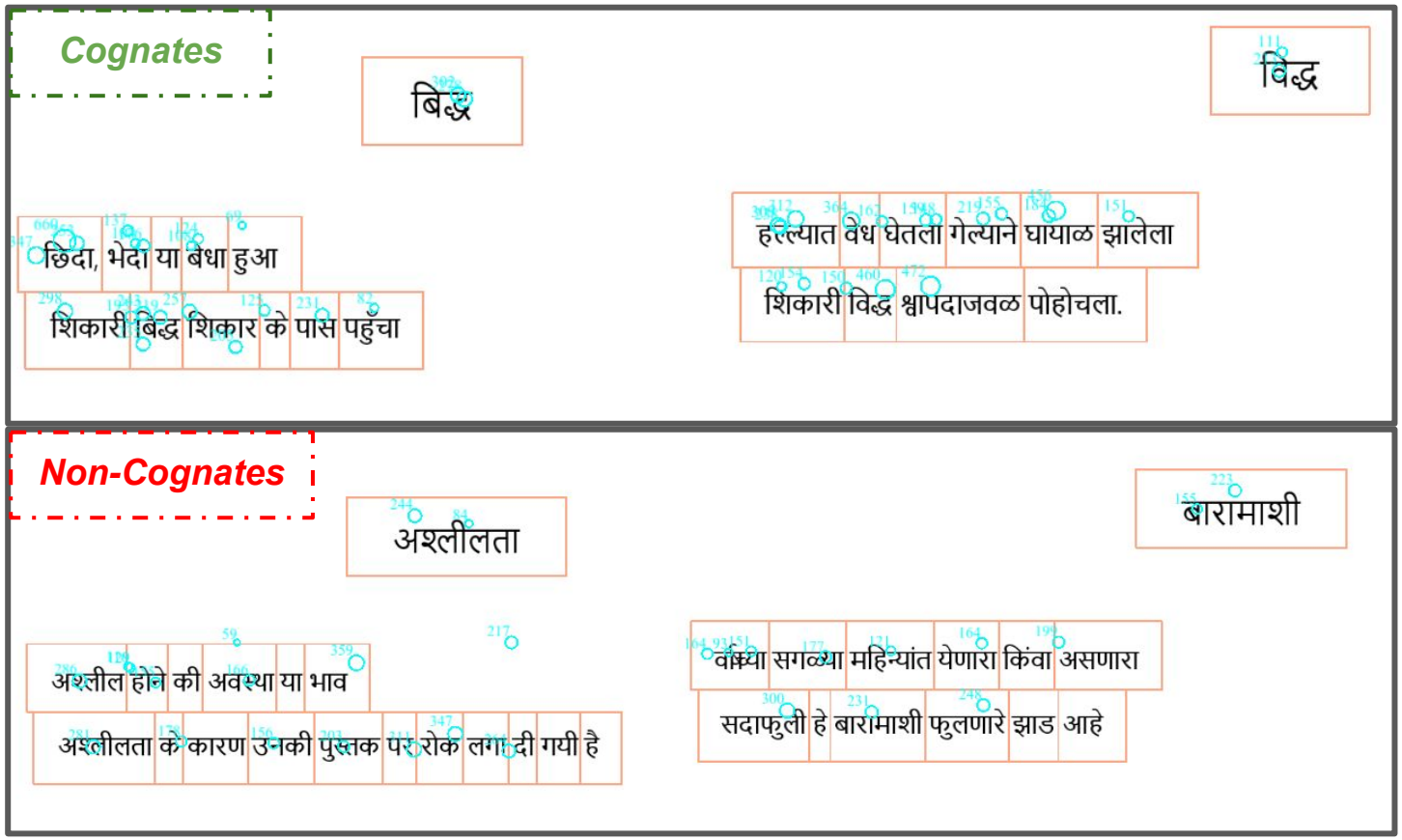
To **verify the annotation quality** we observed two key aspects

- Annotation Precision
  (both individual and aggregate)
- Inter Annotator Agreement among our nine annotators
  (Fleiss Kappa Score)

Dataset Collection

GOAL:

- Given cognate, and non-cognate pair along with their context (definition and example) collect gaze features for two hundred samples (100 +ver, 100 -ve).



Annotator Precision and Inter-annotator Agreement

Annotator	A1	A2	A3	A4	A5	A6	A7	A8	A9	Average
Precision	0.99	0.975	0.965	0.995	0.995	0.99	0.975	0.99	0.98	**0.9839**

Statistical Significance	Value
P-bar	0.005272
P-bar-e	23.7219
Fleiss Kappa	**1.0002**

Cohen's Kappa vs. Fleiss' Kappa

Statistical literature observes that Cohen's kappa is **applicable to two annotators**

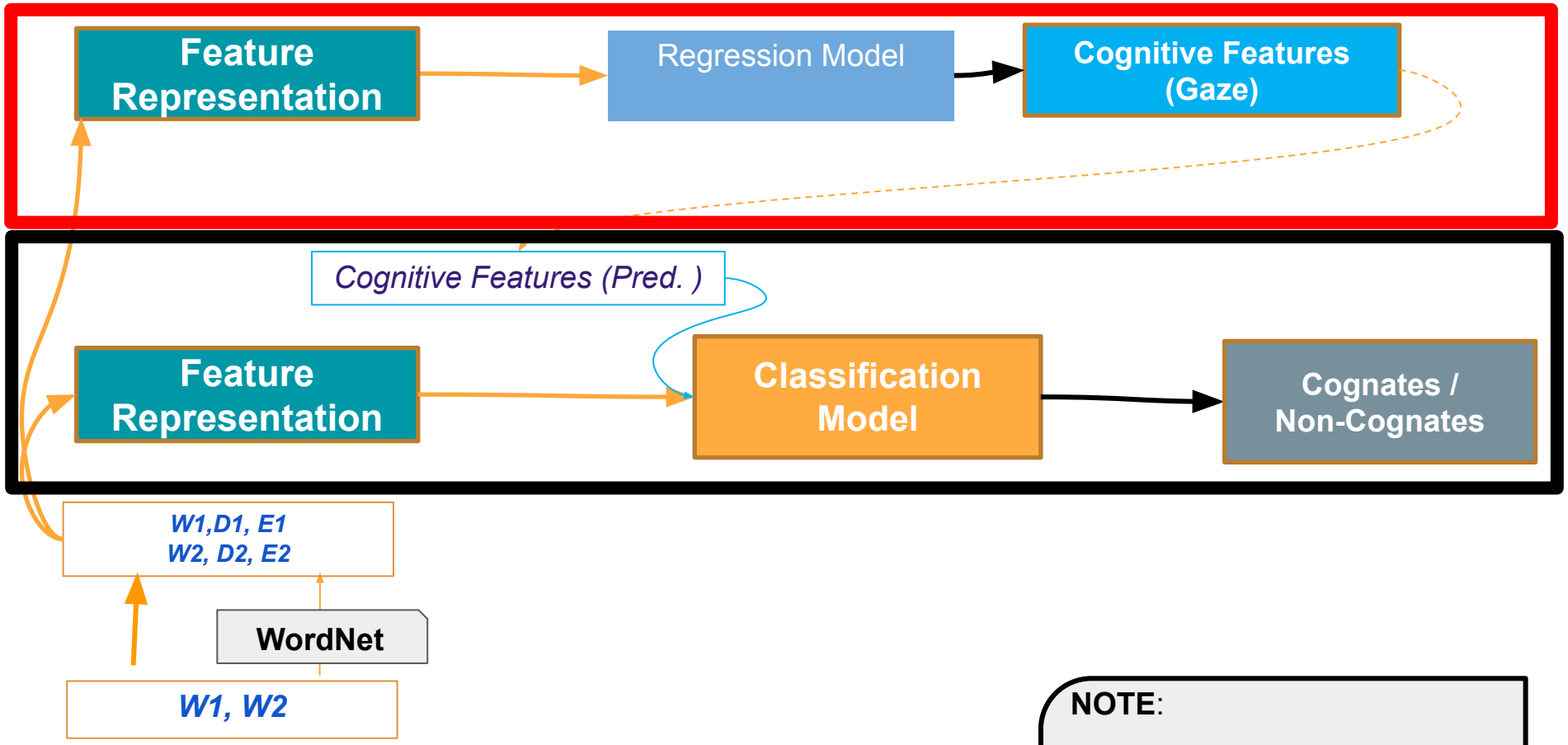
There are studies which use Cohen's kappa for multiple annotators by computing a mean.

Fleiss' kappa, however, **allows multiple annotators**, and categorical values to be taken into account.

We use Fleiss' Kappa for statistical significance.

Cognate Detection with Gaze Features

Proposed Model 1 **: Neural Model for Cognition aware Cognate Detection**



- The regression and classification model are trained separately
- **Two Stages:**
  - Stage-1: Cognitive Feature Prediction
  - Stage-2: Classification model for Cognate Detection

NOTE:

W1, W2: word pairs from two languages

D1, D2: definition of w1,w2

E1, E2: example of w1, w2

Results

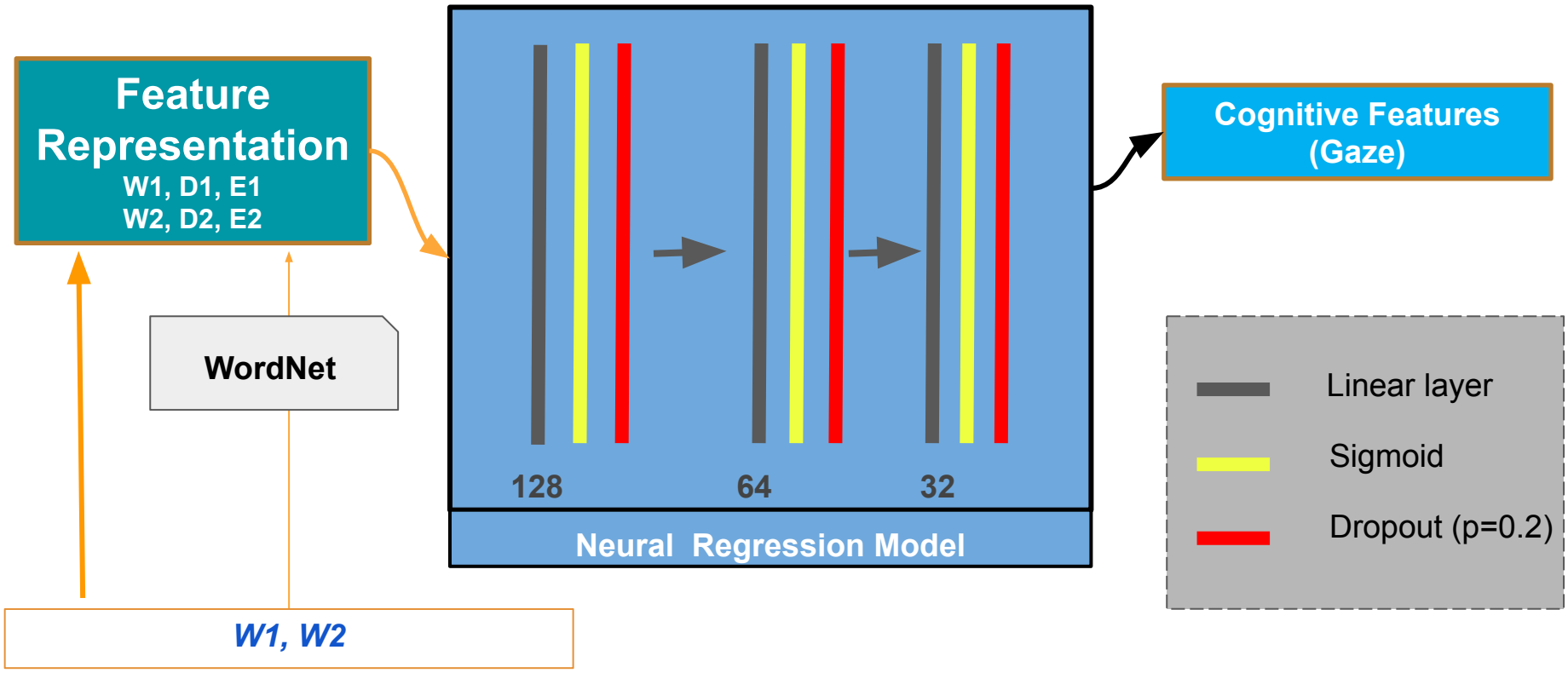
	P	R	F	P	R	F	P	R	F	P	R	F
Feature Set →	Phonetic			WLS								
Rama et. al., 2016 (D1+D2)	0.71	0.69	0.70	-	-	-						
Kanojia et. al., 2019 (D1+D2)	-	-	-	0.76	0.72	0.74						
Feature Set →	XLM			MUSE			VecMap					
Linear SVM (D1+D2)	0.83	0.71	0.77	0.72	0.68	0.70	0.70	0.65	0.67			
LogisticRegression (D1+D2)	0.85	0.74	0.79	0.80	0.71	0.75	0.70	0.66	0.68			
FFNN (D1 + D2)	0.82	0.84	**0.83**	0.83	0.79	0.81	0.75	0.76	0.75			
Feature Set →	XLM+Gaze			MUSE+Gaze			VecMap+Gaze			Gaze		
Linear SVM (D2)	0.81	0.69	0.75	0.72	0.73	0.72	0.70	0.75	0.72	0.77	0.76	0.76
LogisticRegression (D2)	0.84	0.75	0.79	0.76	0.72	0.74	0.81	0.71	0.76	0.80	0.75	0.77
FFNN (D2)	0.83	0.85	**0.84**	0.83	0.78	0.80	0.86	0.83	0.84	0.81	0.71	0.76
Predicted Gaze Features On D1 (11652 samples) and Collected Gaze Features on D2 (200 samples)												
Feature Set →	XLM+Gaze			MUSE+Gaze			VecMap+Gaze			Gaze		
FFNN (D1 + D2)	0.84	0.88	**0.86**	0.85	0.78	0.81	0.83	0.85	0.84	0.77	0.76	0.76
FFNN (D1) [Only Predicted Gaze]	0.83	0.84	0.83	0.82	0.79	0.80	0.80	0.86	0.83	0.76	0.77	0.76

Model 1: Observation

- Our experiments shows that Introducing Gaze Features, results in improving cognate detection accuracy.
- Even on limited samples (1800 samples), our model shows improvement for the task of cognate detection
- Leveraging context information using neural architecture can help improving cognate detection accuracy.

Cognitive Features Prediction

Proposed Model 2 : Neural Model for Cognitive Feature prediction



- **Single Stage**
- MSE loss

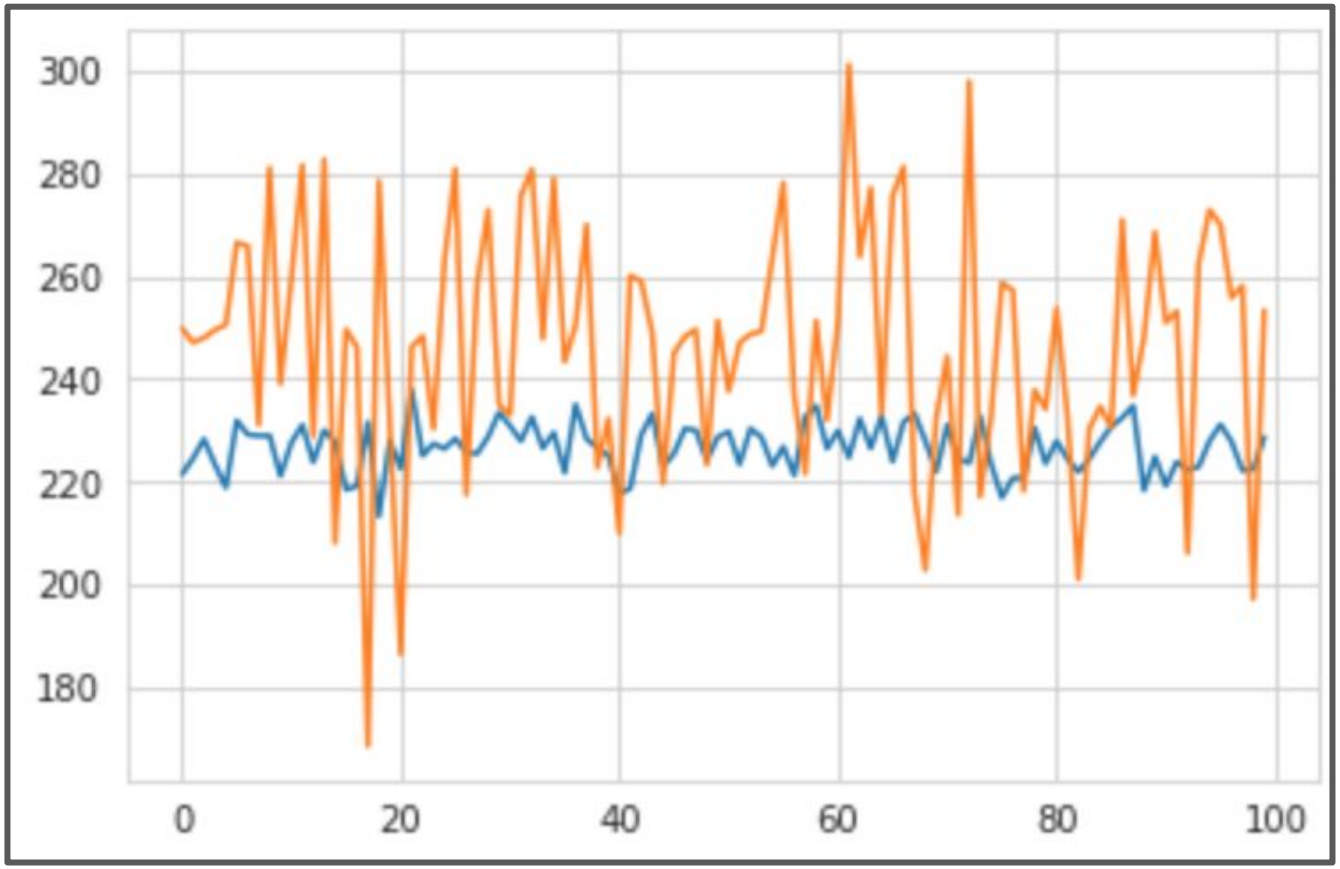
NOTE:

W1, W2: word pairs from two languages

D1, D2: definition of w1,w2

E1, E2: example of w1, w2

Model 1: Results

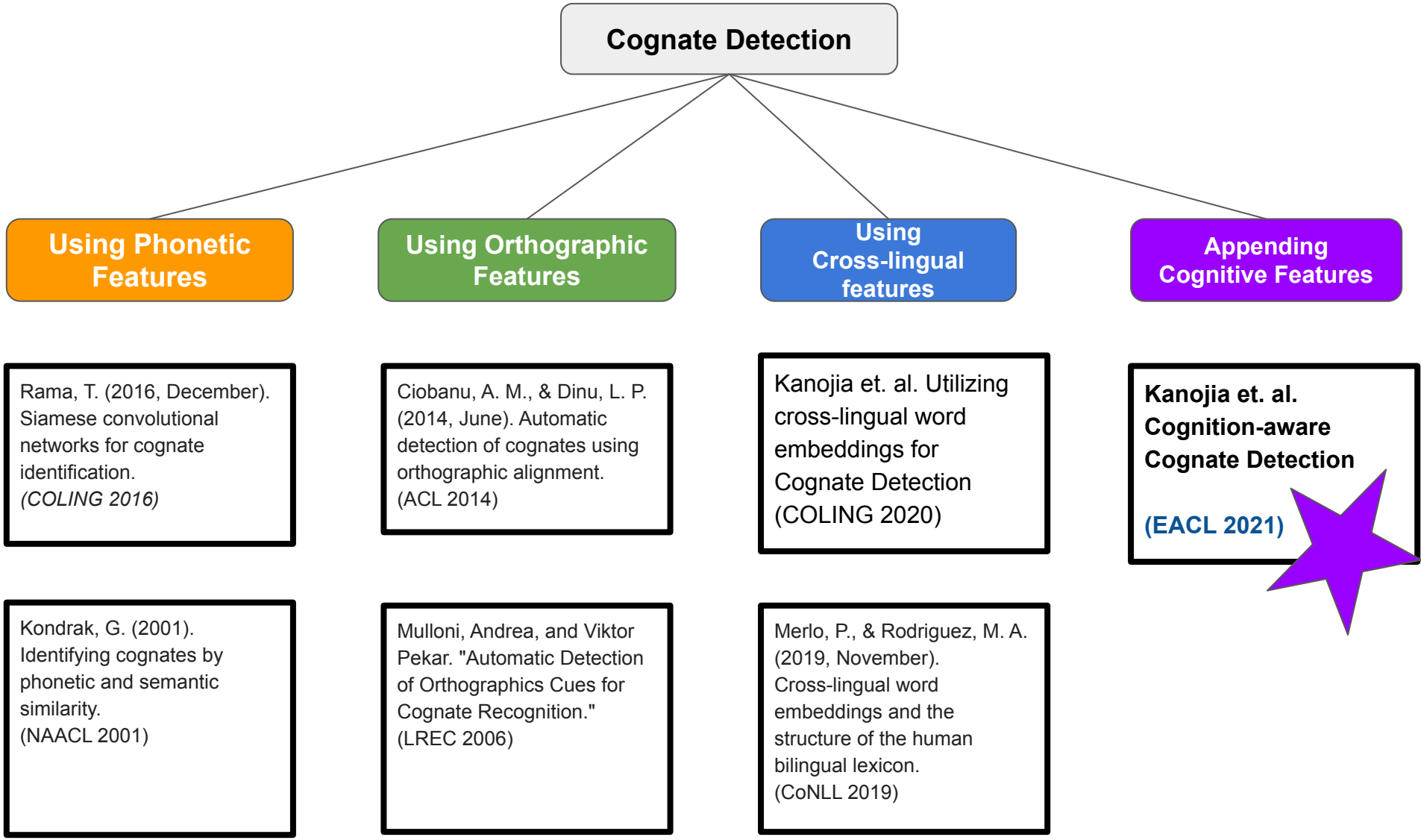


1. AVERAGE FIXATION DURATION
2. AVERAGE SACCADE AMPLITUDE
3. FIXATION COUNT
4. FIXATION DURATION MAX
5. FIXATION DURATION MIN
6. IA COUNT
7. RUN COUNT
8. SACCADE COUNT

Results

	P	R	F	P	R	F	P	R	F	P	R	F
Feature Set →	Phonetic			WLS								
Rama et. al., 2016 (D1+D2)	0.71	0.69	0.70	-	-	-						
Kanojia et. al., 2019 (D1+D2)	-	-	-	0.76	0.72	0.74						
Feature Set →	XLM			MUSE			VecMap					
Linear SVM (D1+D2)	0.83	0.71	0.77	0.72	0.68	0.70	0.70	0.65	0.67			
LogisticRegression (D1+D2)	0.85	0.74	0.79	0.80	0.71	0.75	0.70	0.66	0.68			
FFNN (D1 + D2)	0.82	0.84	**0.83**	0.83	0.79	0.81	0.75	0.76	0.75			
Feature Set →	XLM+Gaze			MUSE+Gaze			VecMap+Gaze			Gaze		
Linear SVM (D2)	0.81	0.69	0.75	0.72	0.73	0.72	0.70	0.75	0.72	0.77	0.76	0.76
LogisticRegression (D2)	0.84	0.75	0.79	0.76	0.72	0.74	0.81	0.71	0.76	0.80	0.75	0.77
FFNN (D2)	0.83	0.85	**0.84**	0.83	0.78	0.80	0.86	0.83	0.84	0.81	0.71	0.76
Predicted Gaze Features On D1 (11652 samples) and Collected Gaze Features on D2 (200 samples)												
Feature Set →	XLM+Gaze			MUSE+Gaze			VecMap+Gaze			Gaze		
FFNN (D1 + D2)	0.84	0.88	**0.86**	0.85	0.78	0.81	0.83	0.85	0.84	0.77	0.76	0.76
FFNN (D1) [Only Predicted Gaze]	0.83	0.84	0.83	0.82	0.79	0.80	0.80	0.86	0.83	0.76	0.77	0.76

Future Investigation



Future Work

- Cognate Detection using predicted gaze features on full corpus (Kanojia et. al. 2020)
- Multi-Task Learning to predict gaze features and use it to predict a label for whether the words are cognates or not.
- Leveraging richer context representation for task of cognate detection.
- Predicting cognitive features for major NLP tasks: eg. Sentiment Analysis, Sarcasm Detection etc.
- Leveraging cognitive features for the task of word sense disambiguation.

References

- Frunza, Oana, and Diana Inkpen. "Semi-supervised learning of partial cognates using bilingual bootstrapping." Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics. 2006.
- Sáchez-Casas, Rosa M., José E. García-Albea, and Christopher W. Davis. "Bilingual lexical processing: Exploring the cognate/non-cognate distinction." European Journal of Cognitive Psychology 4.4 (1992): 293-310.
- Rama, Taraka, et al. "Are automatic methods for cognate detection good enough for phylogenetic reconstruction in historical linguistics?." arXiv preprint arXiv:1804.05416 (2018).
- Mulloni, Andrea, and Viktor Pekar. "Automatic Detection of Orthographics Cues for Cognate Recognition." LREC. 2006.
- Ciobanu, Alina Maria, and Liviu P. Dinu. "Automatic detection of cognates using orthographic alignment." Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). 2014.
- Kondrak, Grzegorz. "Identifying cognates by phonetic and semantic similarity." Second Meeting of the North American Chapter of the Association for Computational Linguistics. 2001.

Thank You!