People-Centred AI
UNIVERSITY OF SURREY

UNIVERSITY OF SURREY

Cognitive Natural Language Processing
(Cognitive NLP)

Dr Diptesh Kanojia

Surrey Institute for People-Centred AI, Department of Computer Science

(This research was performed at the CFILT Lab, IIT Bombay)

Roadmap

Cognitive Science

Eye-mind Hypothesis

Eye-tracking Infrastructure

Understanding Eye-tracking Output

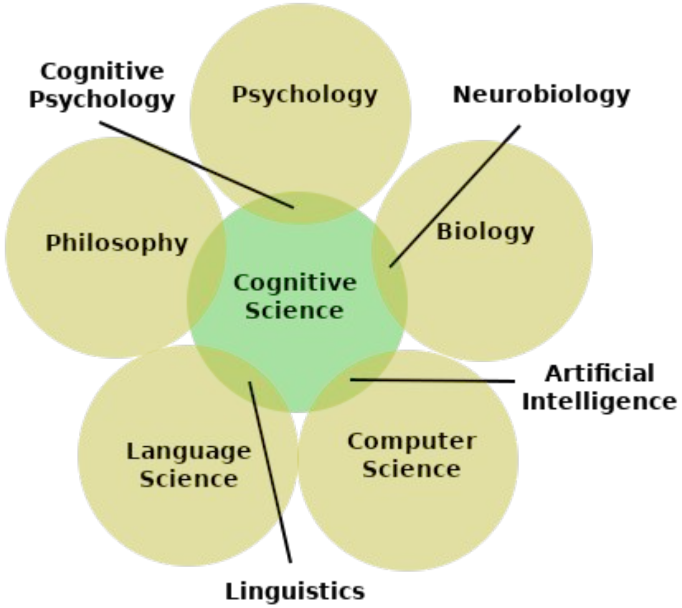
Eye-tracking Features

Utilization in ML/DL pipelines

Cognitive Science

Cognitive science is the interdisciplinary study of mind and intelligence, embracing philosophy, psychology, artificial intelligence, neuroscience, linguistics, and anthropology.

We are interested in is Computational Psycholinguistics and to be more precise, **Cognitive NLP**.

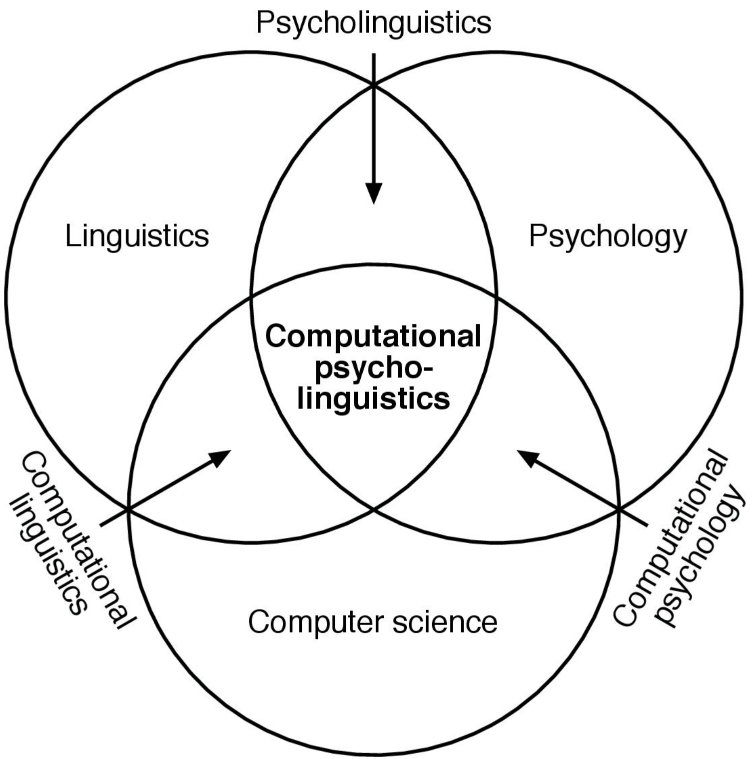


Reading: https://plato.stanford.edu/entries/cognitive-science/

Psycholinguistics and Related Studies

Psycholinguistics is the discipline that investigates and describes the psychological processes that make it possible for humans to master and use language.

- Inferring brain activity using eye movements **(Eye-tracking or Gaze-tracking)**
- Brain activity reading using Electroencephalogram (EEG)
- Brain activity reading using Magnetic Resonance Imaging (MRI)
- However, in the psycholinguistics area - all the above studies are performed on a linguistic theory with reasonable assumptions.



UNIVERSITY OF SURREY

Eye-mind Hypothesis

“the eye remains fixated on a word as long as the word is being processed”

OR

“whatever the eye sees, that is what the mind processes”

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329–354.

However, care needs to be taken in respect of:

- The context of eye behaviour; e.g. a specific search task allows more confidence in inferences drawn whereas an open brief to look, means more factors are likely to influence behaviour, such as meaningfulness, visual (bottom-up) cues and motivational level
- Expectations, experience and individual differences will also influence behaviour
- The role of peripheral vision and pre-attentive processing cannot be directly determined by eye tracking and need to be inferred from eye movement data

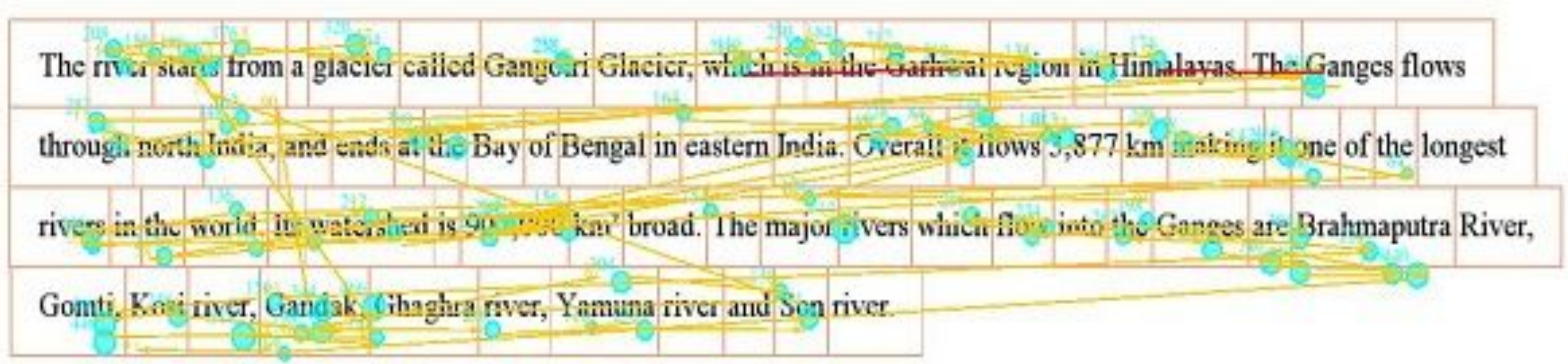
Mindreading From the Eyes Declines With Aging – Evidence From 1,603 Subjects

Jana Kynast[1,2], Eva Maria Quinque[1,2], Maryna Polyakova[1,2], Tobias Luck[3], Steffi G. Riedel-Heller[2,4], Simon Baron-Cohen[5], Andreas Hinz[6], A. Veronica Witte[1,2], Julia Sacher[1,2,7], Arno Villringer[1,2,7] and Matthias L. Schroeter[1,2,7*]

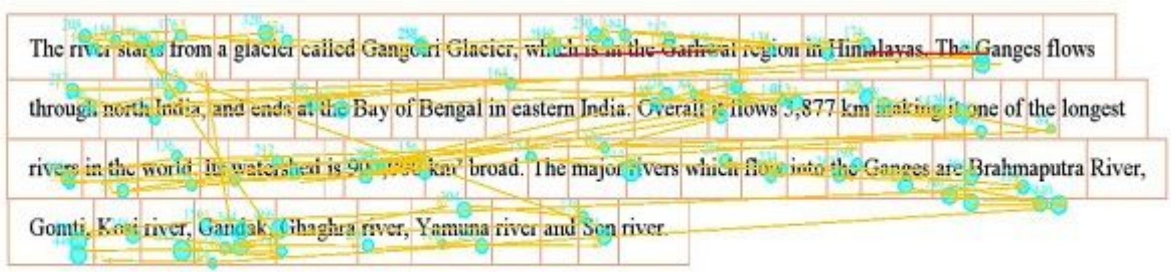
[1]Department of Neurology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
[2]LIFE – Leipzig Research Center for Civilization Diseases, University of Leipzig, Leipzig, Germany
[3]Faculty of Applied Social Sciences, University of Applied Sciences Erfurt, Erfurt, Germany
[4]Institute for Social Medicine, Occupational Medicine and Public Health, University Hospital Leipzig, Leipzig, Germany
[5]Department of Psychiatry, Autism Research Centre, University of Cambridge, Cambridge, United Kingdom
[6]Department for Medical Psychology and Sociology, University Hospital Leipzig, Leipzig, Germany
[7]Clinic for Cognitive Neurology, University Hospital Leipzig, Leipzig, Germany

Infrastructure

Eye Tracking Output



Understanding Eye-tracking output



Interest Area (IA): An interest area (IA) is the area of the screen that is of interest.

Fixation(s): A fixation is an event that takes place when the eye is focused on a point of the screen. That point could either be an interest area, or the screen's background

Saccade(s): A saccade is the rapid movement of the eye from one fixation point to the next. There are two types of saccades - **regressions** and **progressions** . **Regressions** take place when the eye moves from the current interest area to an earlier one. **Progressions** take place when the eye moves from the current interest area to a later one.

Basic Feature Set

1. **AVERAGE_FIXATION_DURATION:**
   a. the average of all fixation duration across all interest areas
2. **AVERAGE_SACCADE_AMPLITUDE:**
   a. saccade amplitude is amplitude of going back and forth - duration
3. **FIXATION_COUNT**
4. **FIXATION_DURATION_MAX:** max time for a single fixation on any IA
5. **FIXATION_DURATION_MIN:** min time for fixation on any IA
   a. User can look at one word
6. **IA_COUNT:** Interest area count
   a. Let's say the context contains relatively higher no. of words.
7. **RUN_COUNT**
   a. Consecutive counts for same IA are ignored in the Run Count
8. **SACCADE_COUNT**
   a. Total count of saccades - user may go back and forth on a screen b/w two points

Eye-tracking and ML/DL Synergization

As an Illustrative example, let's look at our work on

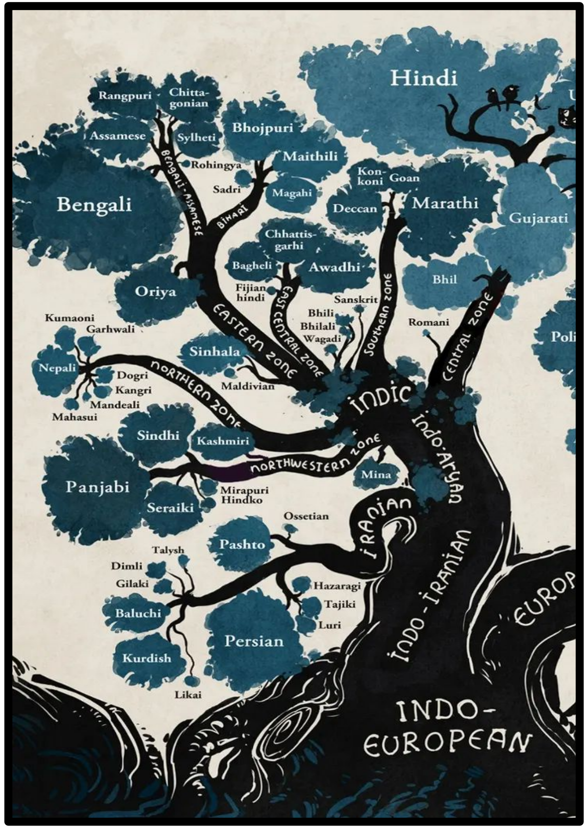
"Cognition-aware Cognate Detection" (EACL-21, Best Long Papers)

We ask the following pertinent questions with this work:

- *Can cognitive features be used to help the task of Cognate Detection*?
- Additionally, Using gaze features collected on a small set of data points, *can we predict the same features on a larger set of data points* to alleviate the need for collecting gaze data?

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

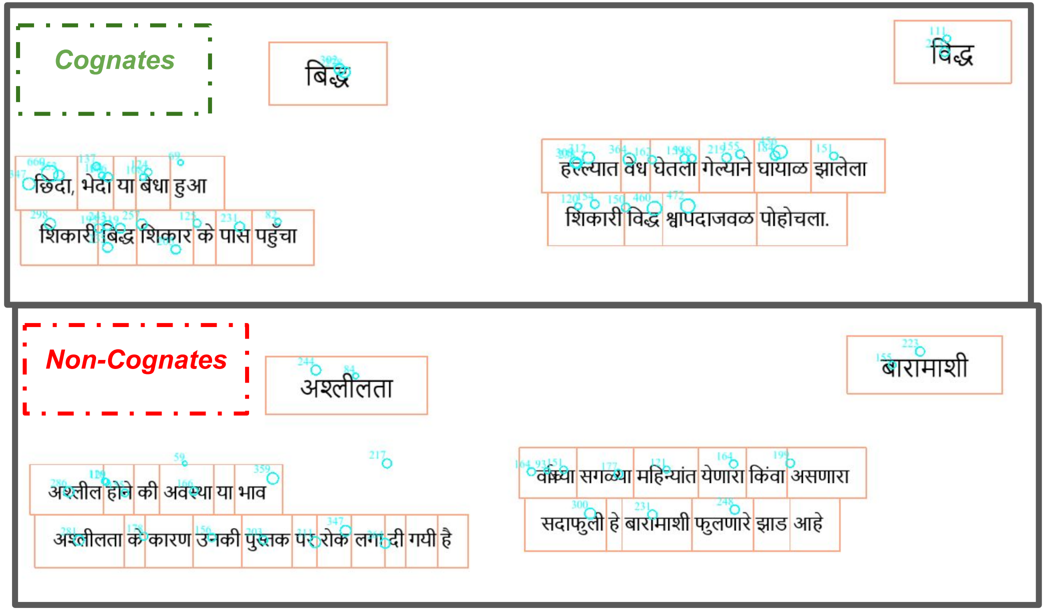
OUTPUT

Cognates (1) /
Non-Cognates (0)

Dataset Collection

GOAL:

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



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)

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

Dataset Collection Setup

Annotator Info

- Nine annotators
- Bilingual Native Marathi speakers

  (who understand Hindi)
- 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)

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. Flei

Statistical literature obse **two annotators**

There are studies which use Coh mean.

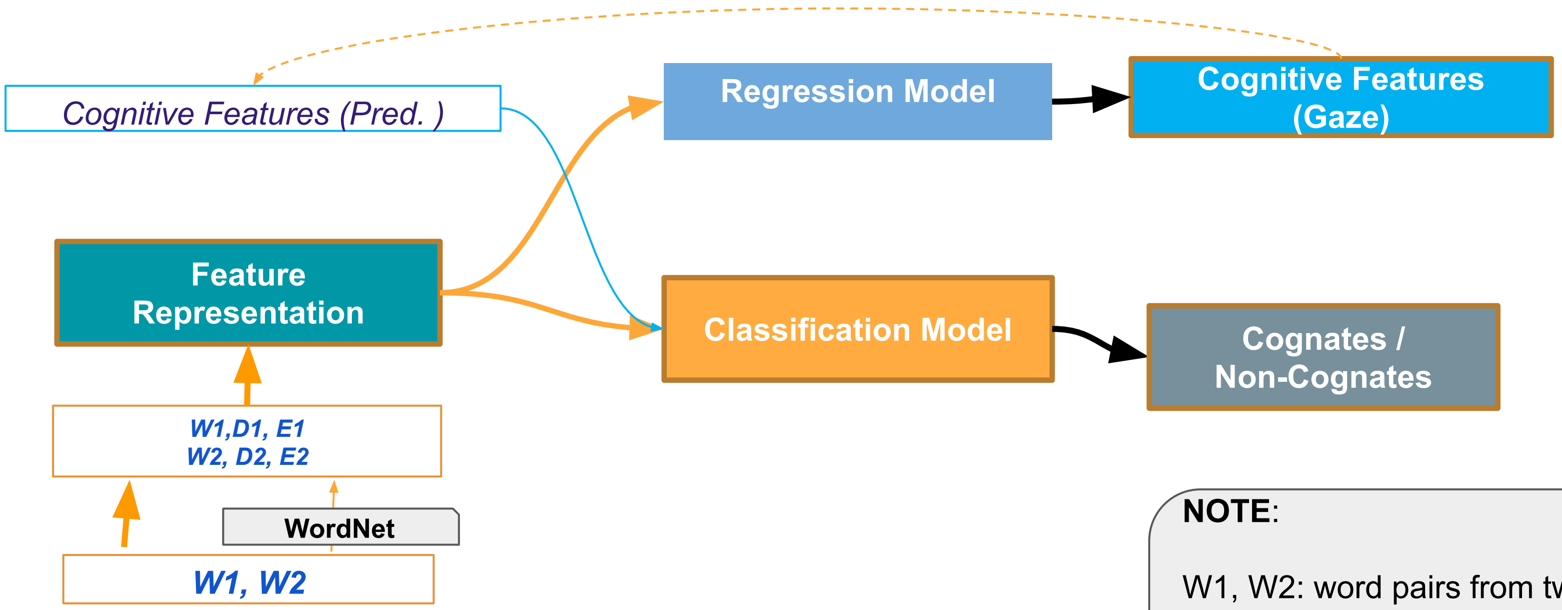
Fleiss' kappa, however, a categorical values to be t

We use Fleiss' Kappa for

	μ_Pos	σ_Pos	μ_Neg	σ_Neg	p
P1	9.720	17.867	8.677	4.281	0.028
P2	8.596	10.526	7.619	13.794	0.049
P3	7.770	6.664	7.044	3.900	0.027
P4	9.686	17.729	8.664	4.306	0.031
P5	8.861	8.611	8.099	5.246	0.042
P6	7.854	6.286	7.184	3.442	0.033
P7	8.564	5.499	7.918	3.540	0.033
P8	8.018	5.955	7.340	3.742	0.031
P9	9.720	17.867	8.703	4.305	0.028

Table 3: T-test statistics for average fixation duration time per word for Positive labels (Cognates) and Negative labels (False Friends) for participants P1-P9.

Proposed Model 1 : Neural Model for Cognition aware Cognate Detection



- **Single Stage**
- The regression and classification model can be trained in Multi-task setting

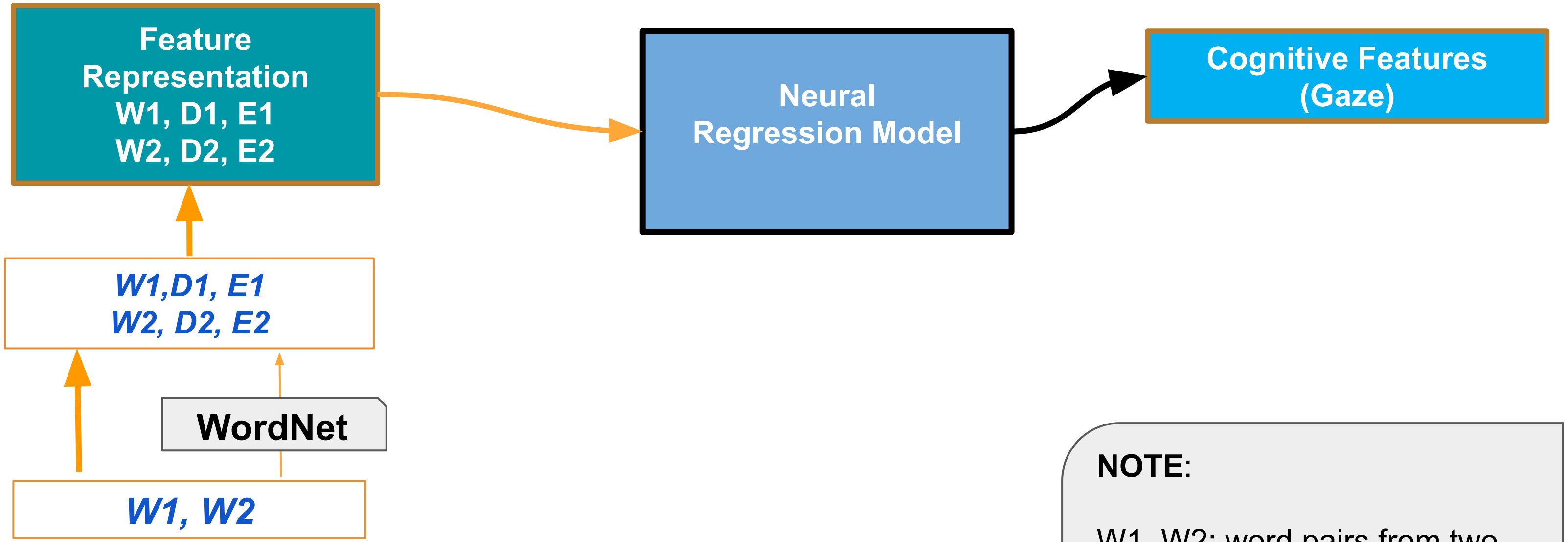
NOTE:

W1, W2: word pairs from two languages

D1, D2: definition of w1,w2

E1, E2: example of w1, w2

Proposed Model 2 : Neural Model for Cognitive Feature prediction



- **Single Stage**
- The regression and classification model can be trained in Multi-task setting

NOTE:

W1, W2: word pairs from two languages

D1, D2: definition of w1,w2

E1, E2: example of w1, w2

Gaze Features

1. **AVERAGE_FIXATION_DURATION:**
   a. the average of all fixation duration across all interest areas
2. **AVERAGE_SACCADE_AMPLITUDE:**
   a. saccade amplitude is amplitude of going back and forth - duration
3. **FIXATION_COUNT**
4. **FIXATION_DURATION_MAX:** max time for a single fixation on any IA
5. **FIXATION_DURATION_MIN:** min time for fixation on any IA
   a. User can look at one word
6. **IA_COUNT:** Interest area count
   a. Let’s say the context contains relatively higher no. of words.
7. **RUN_COUNT**
   a. Consecutive counts for same IA are ignored in the Run Count
8. **SACCADE_COUNT**
   a. Total count of saccades - user may go back and forth on a screen b/w two points

	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

Table 5: Classification results in terms of weighted Precision (P), Recall (R), and F-scores (F) using 5-fold cross-validation using different feature sets as described above.

Observation based on Model 1

- 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 Feature Prediction

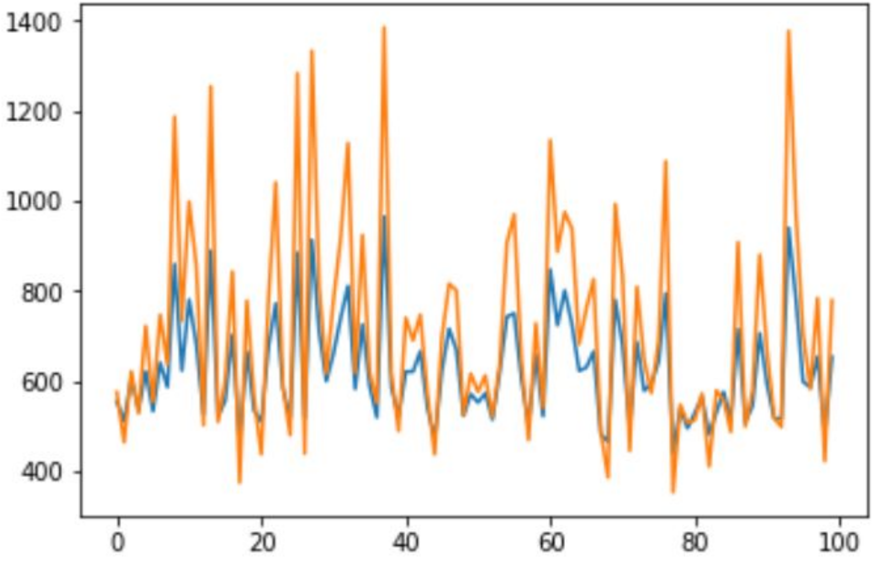


Figure 2: Predicted feature values (blue) vs. Gold feature values (orange) for the average fixation duration feature, on 100 samples.

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'

Observation based on Model 2

- We were able to predict gaze features by learning a neural regression model.
- We also noticed that since the gold gaze features for both cognates and non-cognates are close by.
- Thus, we probably need to introduce a better loss function to take into consideration the distribution for various gaze features for cognates *vs.* non-cognates.
- Alternatively, perform both tasks in a multi-task learning setup.

Conclusion & Future Direction

Interdisciplinary area

Combines gaze tracking with linguistics and computer science.

Gaze features help multiple NLP tasks and detection of cognates is challenging from a traditional NLP perspective.

Quality Estimation for Machine Translation

Regression Task instead of Classification

Does human gaze have a reading pattern when going through well-translated output *vs.* badly-translated output?

Thank you!

Questions?

For offline discussions - d.kanojia@surrey.ac.uk

For papers and resources on existing research:
https://www.cfilt.iitb.ac.in/cognitive-nlp/