Data in Cognitive Science

Frank Keller

Institute for Language, Cognition and Computation
School of Informatics, University of Edinburgh

October 30, 2014

Cognitive Science

The aim of cognitive science is to figure out how the mind works.

Cognitive Science

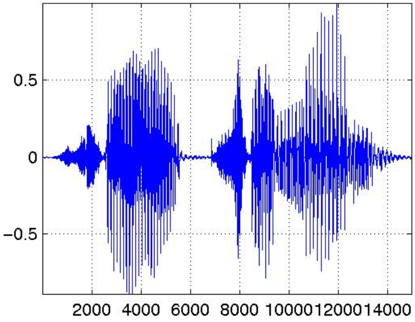
Cognitive scientists do this by studying a range of cognitive processes:

- vision;

Cognitive Science

Cognitive scientists do this by studying a range of cognitive processes:

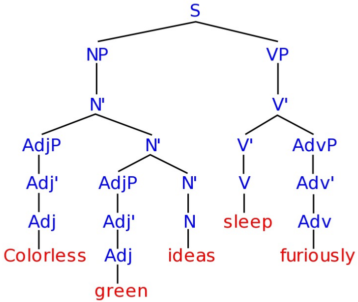
- vision;
- auditory processing;



Cognitive Science

Cognitive scientists do this by studying a range of cognitive processes:

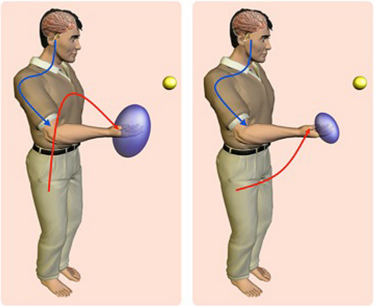
- vision;
- auditory processing;
- language;



Cognitive Science

Cognitive scientists do this by studying a range of cognitive processes:

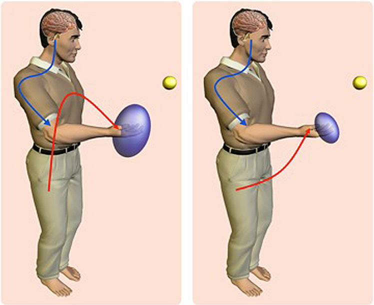
- vision;
- auditory processing;
- language;
- motor control.



Cognitive Science

Cognitive scientists do this by studying a range of cognitive processes:

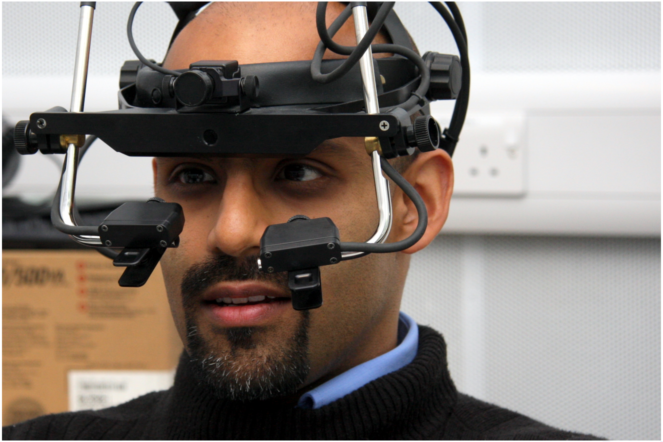
- vision;
- auditory processing;
- language;
- motor control.



Here, we will focus on vision and language.

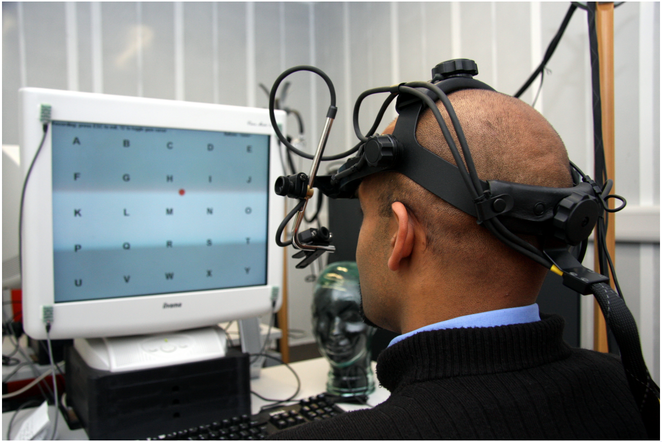
Eye-tracking

By recording their eye-movements, we can study how humans process linguistic or visual information.



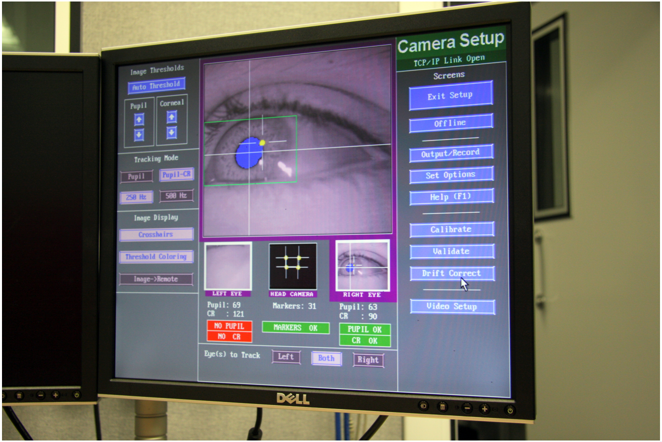
Eye-tracking

By recording their eye-movements, we can study how humans process linguistic or visual information.



Eye-tracking

By recording their eye-movements, we can study how humans process linguistic or visual information.



Visual Attention

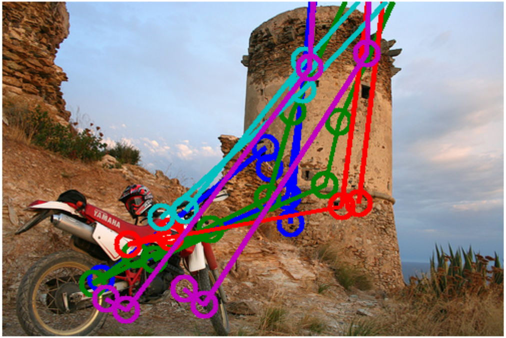
When people view an image, their eyes remains static for short periods of time (fixation), and then jump to a new location (saccade).

Visual Attention

When people view an image, their eyes remains static for short periods of time (fixation), and then jump to a new location (saccade).

Visual Attention

When people view an image, their eyes remains static for short periods of time (fixation), and then jump to a new location (saccade).



Visual Attention

When people view an image, their eyes remains static for short periods of time (fixation), and then jump to a new location (saccade).

Visual Attention

When people view an image, their eyes remains static for short periods of time (fixation), and then jump to a new location (saccade).

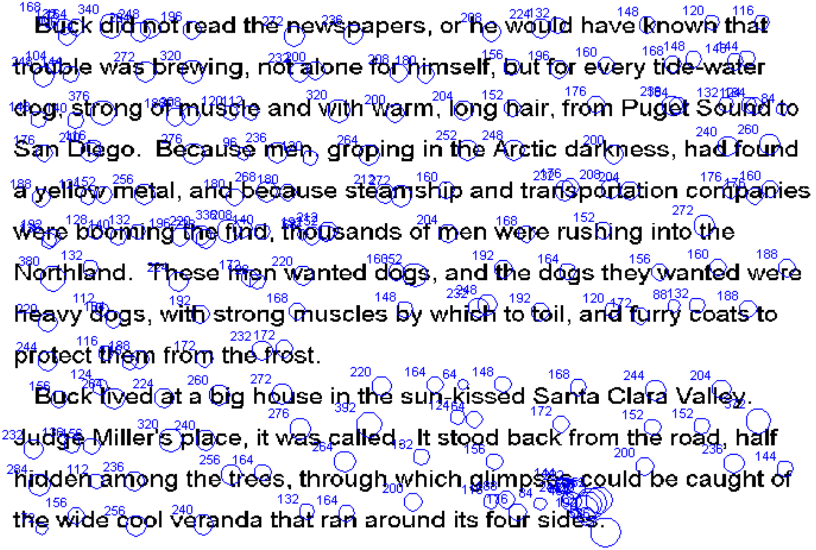
More fixations on salient regions, out-of-context objects, animate objects, task-relevant areas.

Language Processing

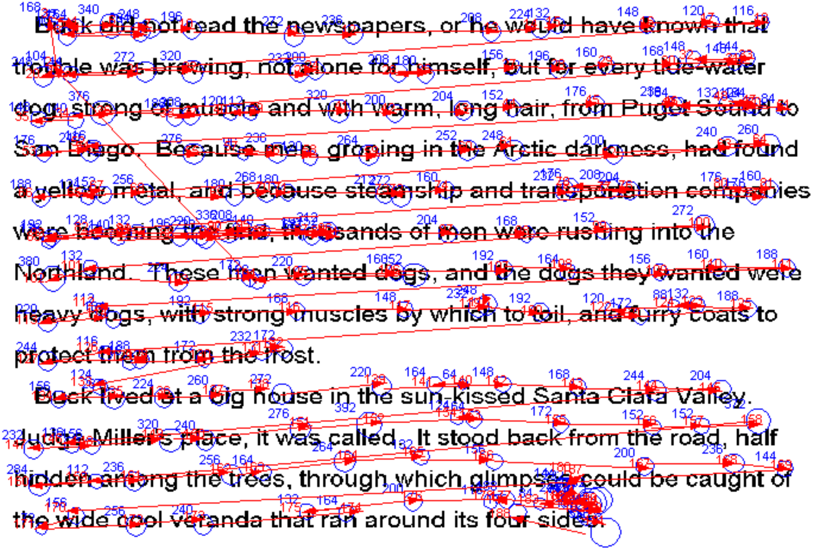
Buck did not read the newspapers, or he would have known that trouble was brewing, not alone for himself, but for every tide-water dog, strong of muscle and with warm, long hair, from Puget Sound to San Diego. Because men, groping in the Arctic darkness, had found a yellow metal, and because steamship and transportation companies were booming the find, thousands of men were rushing into the Northland. These men wanted dogs, and the dogs they wanted were heavy dogs, with strong muscles by which to toil, and furry coats to protect them from the frost.

Buck lived at a big house in the sun-kissed Santa Clara Valley. Judge Miller's place, it was called. It stood back from the road, half hidden among the trees, through which glimpses could be caught of the wide cool veranda that ran around its four sides.

Language Processing



Language Processing



Language Processing

When people read text, they also make fixations and saccades:

- words that are longer or more frequent are fixated for longer;
- syntactic ambiguity leads to re-reading (reverse saccades);
- fixation duration also varies with semantic plausibility, sentence and discourse context, reading task.

Language Processing

When people read text, they also make fixations and saccades:

- words that are longer or more frequent are fixated for longer;
- syntactic ambiguity leads to re-reading (reverse saccades);
- fixation duration also varies with semantic plausibility, sentence and discourse context, reading task.

We can use eye-tracking data to build probabilistic parsing models that simulate human language processing.

Language Processing

When people read text, they also make fixations and saccades:

- words that are longer or more frequent are fixated for longer;
- syntactic ambiguity leads to re-reading (reverse saccades);
- fixation duration also varies with semantic plausibility, sentence and discourse context, reading task.

We can use eye-tracking data to build probabilistic parsing models that simulate human language processing.

Human parsing is word-by-word incremental. Example:

Language Processing

When people read text, they also make fixations and saccades:

- words that are longer or more frequent are fixated for longer;
- syntactic ambiguity leads to re-reading (reverse saccades);
- fixation duration also varies with semantic plausibility, sentence and discourse context, reading task.

We can use eye-tracking data to build probabilistic parsing models that simulate human language processing.

Human parsing is word-by-word incremental. Example:

(1) Banks

Language Processing

When people read text, they also make fixations and saccades:

- words that are longer or more frequent are fixated for longer;
- syntactic ambiguity leads to re-reading (reverse saccades);
- fixation duration also varies with semantic plausibility, sentence and discourse context, reading task.

We can use eye-tracking data to build probabilistic parsing models that simulate human language processing.

Human parsing is word-by-word incremental. Example:

(1) Banks will

Language Processing

When people read text, they also make fixations and saccades:

- words that are longer or more frequent are fixated for longer;
- syntactic ambiguity leads to re-reading (reverse saccades);
- fixation duration also varies with semantic plausibility, sentence and discourse context, reading task.

We can use eye-tracking data to build probabilistic parsing models that simulate human language processing.

Human parsing is word-by-word incremental. Example:

(1) Banks will take

Language Processing

When people read text, they also make fixations and saccades:

- words that are longer or more frequent are fixated for longer;
- syntactic ambiguity leads to re-reading (reverse saccades);
- fixation duration also varies with semantic plausibility, sentence and discourse context, reading task.

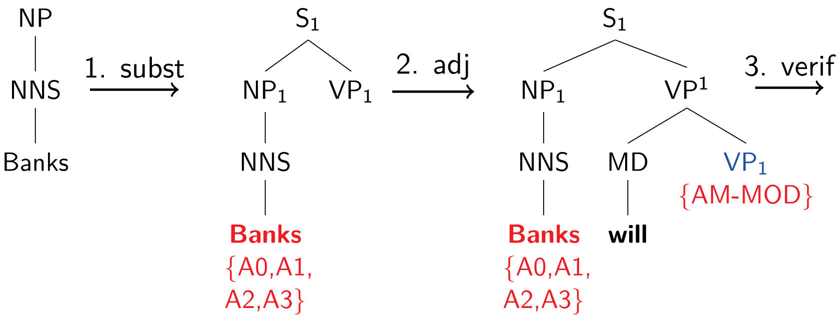
We can use eye-tracking data to build probabilistic parsing models that simulate human language processing.

Human parsing is word-by-word incremental. Example:

(1) Banks will take measures.

Human Parsing

Incremental TAG model proposed by Demberg et al. (2013):

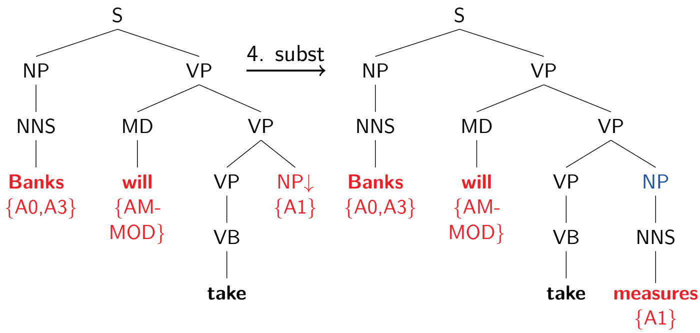


1. NP $\rightarrow \langle$\{A0,A1,A2,A3\},Banks,nil$\rangle$

2. VP $\rightarrow \langle$AM-MOD,will,nil$\rangle$

Human Parsing

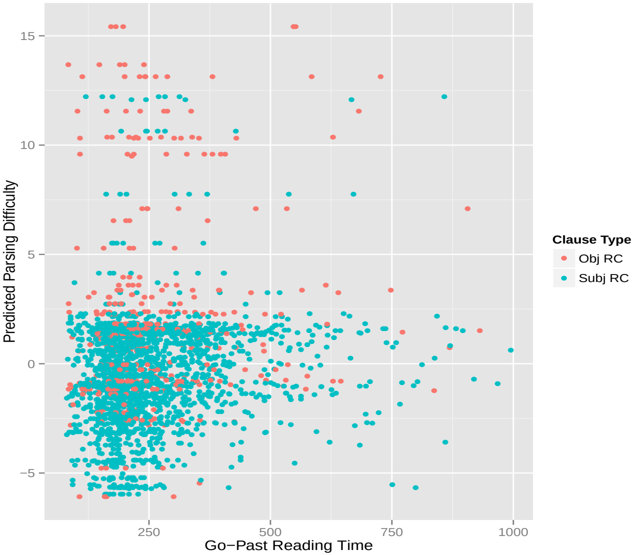
Incremental TAG model proposed by Demberg et al. (2013):



3. NP $\rightarrow$ $\langle$\{A0,A3\},Banks,take$\rangle$
   VP $\rightarrow$ $\langle$AM-MOD,will,take$\rangle$

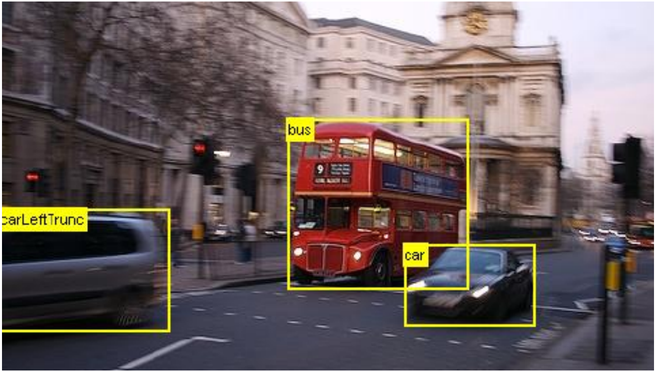
4. NP $\rightarrow$ $\langle$A1,measures,take$\rangle$

Predicting Reading Times



Object Detection

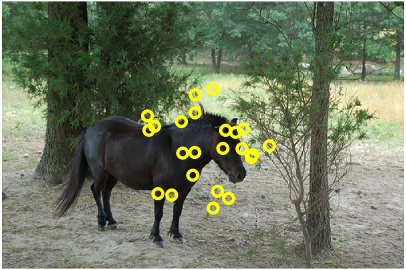
We can use eye-tracking data for a classic computer vision task: object class detection.



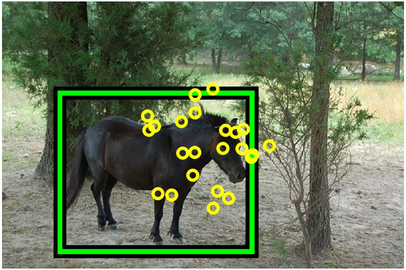
Object detectors are trained on images that are manually annotated with bounding boxed around the objects.

Object Detection

Alternative: infer bounding boxes from eye-tracking data (Papadopoulos et al. 2014):

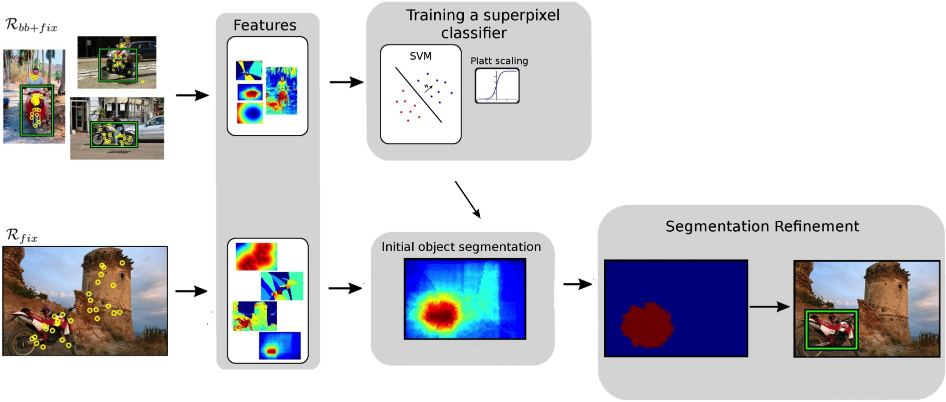


$\Rightarrow$



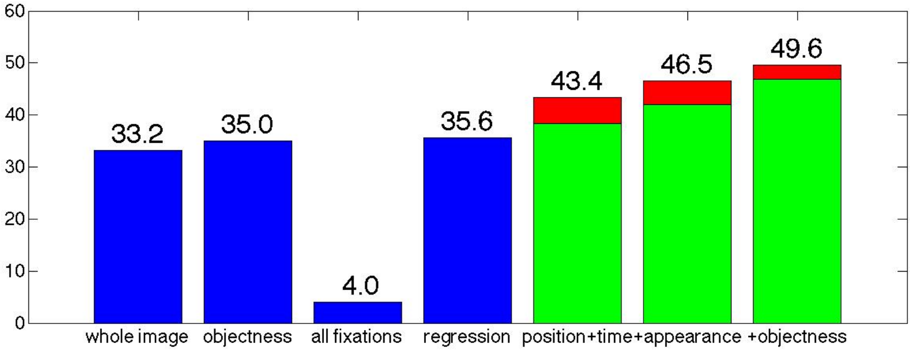
- much faster: 1 s per image vs. 22 s for BB drawing;
- no need for trained annotators, guidelines, etc.

From Fixations to Bounding Boxes



From Fixations to Bounding Boxes

Evaluation using CorLoc (intersection over union > 0.5):



Other Topics in Cognitive Data Science

- Build an incremental semantic role labeler which can be evaluated against eye-tracking data (Konstas et al. 2014);
- use synchronous grammars to align image structure with linguistic structure, e.g., for image description (Elliott & Keller 2013);
- build sequence models (e.g., based on HMMs) that human predict fixation behavior, either for images or for text;
- use eye-tracking data to improve unsupervised PoS tagging or parsing models.

References

Demberg, V., Keller, F., & Koller, A. (2013). Incremental, predictive parsing with psycholinguistically motivated tree-adjoining grammar. *Computational Linguistics*, *39*, 1025–1066.

Elliott, D., & Keller, F. (2013). Image description generation from visual dependency representations. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, (pp. 1292–1302), Seattle, WA.

Konstas, I., Keller, F., Demberg, V., & Lapata, M. (2014). Incremental semantic role labeling with Tree Adjoining Grammar. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, (pp. 301–312), Doha.

Papadopoulos, D. P., Clarke, A. D. F., Keller, F., & Ferrari, V. (2014). Training object class detectors from eye tracking data. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (eds.), *Proceedings of the 13th European Conference on Computer Vision*, vol. V, (pp. 361–376), Zurich.