Between-word regressions as part of rational reading

Klinton Bicknell & Roger Levy

UC San Diego

CUNY 2010: New York

Eye movement control

In psycholinguistics, we study a diverse number of topics

- where do people get garden pathed?
- how are anaphoric links established?
- when are certain information sources available?
- ...

Eye movement control

In psycholinguistics, we study a diverse number of topics

- where do people get garden pathed?
- how are anaphoric links established?
- when are certain information sources available?
- ...

Often, test our predictions with eye tracking in reading

- provides a convenient way to localize our effects

Eye movement control

In psycholinguistics, we study a diverse number of topics

- where do people get garden pathed?
- how are anaphoric links established?
- when are certain information sources available?
- ...

Often, test our predictions with eye tracking in reading

- provides a convenient way to localize our effects
- assumptions linking theoretical predictions to eye movements are loose
  - e.g., more regressions or longer reading times at first pass?
- tightening this link will lead to better understanding of current results, and allow asking new questions

Eye movement control

In psycholinguistics, we study a diverse number of topics

- where do people get garden pathed?
- how are anaphoric links established?
- when are certain information sources available?
- ...

Often, test our predictions with eye tracking in reading

- provides a convenient way to localize our effects
- assumptions linking theoretical predictions to eye movements are loose
  - e.g., more regressions or longer reading times at first pass?
- tightening this link will lead to better understanding of current results, and allow asking new questions

We need to understand how real-time comprehension translates to action

Eye movement control II

E-Z Reader (Reichle et al., 2003) and SWIFT (Engbert et al., 2005) can account for:

- frequency/predictability effects
- word length effects
- fixation durations
- word skipping & refixation rates
- landing site distributions

But:

- they don't address how observed eye-movement patterns may be an adaptive response to the goals of reading

Eye movement control II

E-Z Reader (Reichle et al., 2003) and SWIFT (Engbert et al., 2005) can account for:

- frequency/predictability effects
- word length effects
- fixation durations
- word skipping & refixation rates
- landing site distributions

But:

- they don't address how observed eye-movement patterns may be an adaptive response to the goals of reading

The goal of this talk is to address this issue

A new framework for eye movements in reading

A new framework for eye movements in reading

Readers have a diverse range of goals:

A new framework for eye movements in reading

Readers have a diverse range of goals:

- understanding the point of the passage

A new framework for eye movements in reading

Readers have a diverse range of goals:

- understanding the point of the passage
- going as fast as possible while getting the comprehension question

A new framework for eye movements in reading

Readers have a diverse range of goals:

- understanding the point of the passage
- going as fast as possible while getting the comprehension question

These goals are all accomplished by getting info about the text's identity

A new framework for eye movements in reading

Readers have a diverse range of goals:

- understanding the point of the passage
- going as fast as possible while getting the comprehension question

These goals are all accomplished by getting info about the text's identity

Identifying the text

- two sources of information
  - language knowledge
  - visual input

A new framework for eye movements in reading

Readers have a diverse range of goals:

- understanding the point of the passage
- going as fast as possible while getting the comprehension question

These goals are all accomplished by getting info about the text's identity

Identifying the text

- two sources of information
  - language knowledge (prior)
  - visual input (likelihood)
- normative way to combine these: Bayesian inference

A new framework for eye movements in reading

Readers have a diverse range of goals:

- understanding the point of the passage
- going as fast as possible while getting the comprehension question

These goals are all accomplished by getting info about the text's identity

Identifying the text

- two sources of information
  - language knowledge (prior)
  - visual input (likelihood)
- normative way to combine these: Bayesian inference
- prior is known, so eyes move for visual input

A new framework for eye movements in reading

Readers have a diverse range of goals:

- understanding the point of the passage
- going as fast as possible while getting the comprehension question

These goals are all accomplished by getting info about the text's identity

Identifying the text

- two sources of information
  - language knowledge (prior)
  - visual input (likelihood)
- normative way to combine these: Bayesian inference
- prior is known, so eyes move for visual input

Central hypothesis: eyes move to obtain visual information about text, which helps identify it and thus achieve reader goals

This framework gives a new explanation for regressions

Why make regressions? (moving the eyes backwards)

This framework gives a new explanation for regressions

Why make regressions? (moving the eyes backwards)

- same reason to move the eyes forward: to get visual input

This framework gives a new explanation for regressions

Why make regressions? (moving the eyes backwards)

- same reason to move the eyes forward: to get visual input
- but why get more visual input about previous regions?

This framework gives a new explanation for regressions

Why make regressions? (moving the eyes backwards)

- same reason to move the eyes forward: to get visual input
- but why get more visual input about previous regions?

Confidence about previous regions will sometimes fall

Illustration of confidence about a previous region falling

Illustration of confidence about a previous region falling

'From the closet, she pulled out a *acket ...'

$p(\text{jacket}) \approx .9$
$p(\text{racket}) \approx .1$
$p(\text{packet}) \approx .0$
confidence high

Illustration of confidence about a previous region falling

'From the closet, she pulled out a *acket for the upcoming game ...'

$$
\begin{array}{lcl}
p(\text{jacket}) \approx .9 & \longrightarrow & p(\text{jacket}) \approx .4 \\
p(\text{racket}) \approx .1 & \longrightarrow & p(\text{racket}) \approx .6 \\
p(\text{packet}) \approx .0 & \longrightarrow & p(\text{packet}) \approx .0 \\
\text{confidence high} & \longrightarrow & \text{confidence low}
\end{array}
$$

Illustration of confidence about a previous region falling

'From the closet, she pulled out a *acket for the upcoming game . . . '

$$\begin{array}{lcl} p(\text{jacket}) \approx .9 & \longrightarrow & p(\text{jacket}) \approx .4 \\ p(\text{racket}) \approx .1 & \longrightarrow & p(\text{racket}) \approx .6 \\ p(\text{packet}) \approx .0 & \longrightarrow & p(\text{packet}) \approx .0 \\ \text{confidence high} & \longrightarrow & \text{confidence low} \end{array}$$

Confidence in previous regions will sometimes fall if

- readers use context
- readers maintain uncertainty about previous words (Connine, Blasko, & Hall, 1991; Levy, Bicknell, Slattery, & Rayner, 2009)

Illustration of confidence about a previous region falling

'From the closet, she pulled out a *acket for the upcoming game ...'

$$\begin{array}{lcl} p(\text{jacket}) \approx .9 & \longrightarrow & p(\text{jacket}) \approx .4 \\ p(\text{racket}) \approx .1 & \longrightarrow & p(\text{racket}) \approx .6 \\ p(\text{packet}) \approx .0 & \longrightarrow & p(\text{packet}) \approx .0 \\ \text{confidence high} & \longrightarrow & \text{confidence low} \end{array}$$

Confidence in previous regions will sometimes fall if

- readers use context
- readers maintain uncertainty about previous words (Connine, Blasko, & Hall, 1991; Levy, Bicknell, Slattery, & Rayner, 2009)

Making a regression to *acket seems a reasonable thing to do

This talk

1. present an implemented model of reading within this framework
2. report two simulations with it, demonstrating:
    - regressions are a rational response to confidence falling
    - how model can be flexibly adapted to different reader goals

An implemented model

Framework is general

- start with prior expectations for the text (language knowledge)
- move eyes to get visual input
- update beliefs about text identity as visual input arrives

An implemented model

Framework is general

- start with prior expectations for the text (language knowledge)
- move eyes to get visual input
- update beliefs about text identity as visual input arrives

Pieces of a model in this framework

An implemented model

Framework is general

- start with prior expectations for the text (language knowledge)
- move eyes to get visual input
- update beliefs about text identity as visual input arrives

Pieces of a model in this framework

1. formal problem of reading: possible actions a reader can take

An implemented model

Framework is general

- start with prior expectations for the text (language knowledge)
- move eyes to get visual input
- update beliefs about text identity as visual input arrives

Pieces of a model in this framework

1. formal problem of reading: possible actions a reader can take
2. visual input

An implemented model

Framework is general

- start with prior expectations for the text (language knowledge)
- move eyes to get visual input
- update beliefs about text identity as visual input arrives

Pieces of a model in this framework

1. formal problem of reading: possible actions a reader can take
2. visual input
3. behavior policy: how the model decides between the actions

An implemented model

Framework is general

- start with prior expectations for the text (language knowledge)
- move eyes to get visual input
- update beliefs about text identity as visual input arrives

Pieces of a model in this framework

1. formal problem of reading: possible actions a reader can take
2. visual input
3. behavior policy: how the model decides between the actions

Next up: describing each of these pieces for our current model

The model: Formal reading problem

The model: Formal reading problem

Assume discrete timesteps and on each:

- get visual input about sentence around the fixated character
- update beliefs about sentence identity (vision + context)
- choose an action

The model: Formal reading problem II

Four possible actions

- continue fixating current location
- initiate a forward saccade to position t
- initiate a backward saccade to position t
- stop reading the sentence

The model: Formal reading problem II

Four possible actions

- continue fixating current location (deterministic)
- initiate a forward saccade to position t
- initiate a backward saccade to position t
- stop reading the sentence (deterministic)

The model: Formal reading problem II

Four possible actions

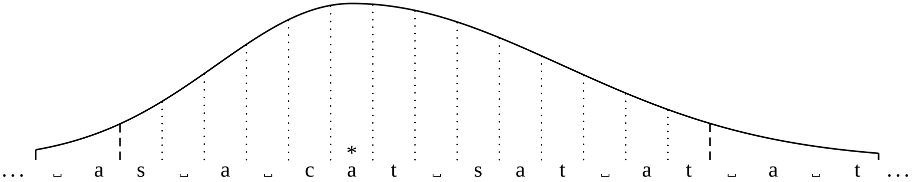
- continue fixating current location (deterministic)
- initiate a forward saccade to position t
- initiate a backward saccade to position t
- stop reading the sentence (deterministic)

When a saccade is initiated

- one timestep of delay (saccade execution lags behind initiation)
- then, landing position normally distributed around t
  - (variance increases with intended distance)

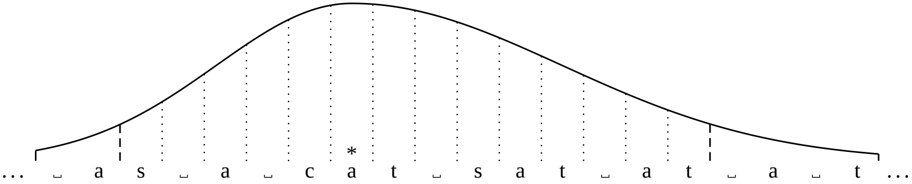
The model: Visual input

The model: Visual input



- acuity curve given by asymmetric Gaussian (as in SWIFT)
- narrow foveal input (14 highest chars)
- wide peripheral input (20 highest chars)

The model: Visual input



Foveal input

- veridical information about word boundaries
- noisy information about letter identity (noise related to acuity)
- no letter confusability (following Norris, 2006)

Peripheral input

- veridical information about word boundaries only

Behavior policy: How to choose actions

Behavior policy: How to choose actions

- model's belief state is a distribution over possible sentences
- decisions made based on *parameterized* policy sensitive to current belief state

Behavior policy: How to choose actions

- model's belief state is a distribution over possible sentences
- decisions made based on *parameterized* policy sensitive to current belief state
- policy might be sensitive to many things
  - expectations about upcoming material
  - ambiguity in structural analysis
  - semantics
  - relevance to expected comprehension question
  - ...

Behavior policy: How to choose actions

- model's belief state is a distribution over possible sentences
- decisions made based on *parameterized* policy sensitive to current belief state
- policy might be sensitive to many things
  - expectations about upcoming material
  - ambiguity in structural analysis
  - semantics
  - relevance to expected comprehension question
  - ...

Today's simple policy

- CONFIDENCE in a character position: probability of most likely character under model beliefs
- move left to right, bringing up confidence in each position to α
- make regression if confidence about a previous position falls below β
  - (if $\beta = 0$, never make a regression)

How to deal with the problem of confidence falling

How to deal with the problem of confidence falling

From the closet, she pulled out a *acket for the upcoming game . . .

How to deal with the problem of confidence falling

From the closet, she pulled out a *acket for the upcoming game . . .

Two strategies

How to deal with the problem of confidence falling

From the closet, she pulled out a *acket for the upcoming game . . .

Two strategies

1. Make regressions when confidence falls

How to deal with the problem of confidence falling

From the closet, she pulled out a *acket for the upcoming game . . .

Two strategies

1. Make regressions when confidence falls
2. Prevent confidence from falling: slow down overall
   - higher confidence earlier $\rightarrow$ confidence unlikely to fall as far

How to deal with the problem of confidence falling

From the closet, she pulled out a *acket for the upcoming game . . .

Two strategies

1. Make regressions when confidence falls
2. Prevent confidence from falling: slow down overall
    - higher confidence earlier → confidence unlikely to fall as far

Intuitively, Strategy 1 may be better

- confidence only falls *sometimes*
- so why slow down across the board?

How to deal with the problem of confidence falling

From the closet, she pulled out a *acket for the upcoming game . . .

Two strategies

1. Make regressions when confidence falls
2. Prevent confidence from falling: slow down overall
    - higher confidence earlier → confidence unlikely to fall as far

Intuitively, Strategy 1 may be better

- confidence only falls *sometimes*
- so why slow down across the board?

Simulation 1: compare performance of these strategies in our model

Formalizing the two strategies

Formalizing the two strategies

- move left to right, bringing up confidence in each position to α
- make regression if confidence about a previous position falls below β

Formalizing the two strategies

- move left to right, bringing up confidence in each position to α
- make regression if confidence about a previous position falls below β

Strategy 1. Make regressions when confidence falls

- relatively lower α
- $\beta > 0$

Formalizing the two strategies

- move left to right, bringing up confidence in each position to α
- make regression if confidence about a previous position falls below β

Strategy 1. Make regressions when confidence falls

- relatively lower α
- $\beta > 0$

Strategy 2. Prevent confidence from falling: slow down overall

- relatively higher α
- $\beta = 0$

Simulation 1

Motivation

- hypothesis: for each nonregressive policy, there is a regressive policy that is faster and more accurate

Simulation 1

Motivation

- hypothesis: for each nonregressive policy, there is a regressive policy that is faster and more accurate
- nonregressive policies tested: $\alpha \in \{.90, .95, .97, .99\}$ and $\beta = 0$

Simulation 1

Motivation

- hypothesis: for each nonregressive policy, there is a regressive policy that is faster and more accurate
- nonregressive policies tested: $\alpha \in \{.90, .95, .97, .99\}$ and $\beta = 0$
- regressive policies tested: $\alpha \in \{.85, .90, .95, .97\}$ and $\beta \in \{.4, .7\}$

Simulation 1

Motivation

- hypothesis: for each nonregressive policy, there is a regressive policy that is faster and more accurate
- nonregressive policies tested: $\alpha \in \{.90, .95, .97, .99\}$ and $\beta = 0$
- regressive policies tested: $\alpha \in \{.85, .90, .95, .97\}$ and $\beta \in \{.4, .7\}$

Measures

1. number of timesteps T until model ends reading
2. 'accuracy' L: log probability of correct sentence under model beliefs

Simulation 1

Motivation

- hypothesis: for each nonregressive policy, there is a regressive policy that is faster and more accurate
- nonregressive policies tested: $\alpha \in \{.90, .95, .97, .99\}$ and $\beta = 0$
- regressive policies tested: $\alpha \in \{.85, .90, .95, .97\}$ and $\beta \in \{.4, .7\}$

Measures

1. number of timesteps T until model ends reading
2. 'accuracy' L: log probability of correct sentence under model beliefs
   - 'how likely is model to guess the sentence correctly?'

Simulation 1

Motivation

- hypothesis: for each nonregressive policy, there is a regressive policy that is faster and more accurate
- nonregressive policies tested: $\alpha \in \{.90, .95, .97, .99\}$ and $\beta = 0$
- regressive policies tested: $\alpha \in \{.85, .90, .95, .97\}$ and $\beta \in \{.4, .7\}$

Measures

1. number of timesteps T until model ends reading
2. 'accuracy' L: log probability of correct sentence under model beliefs
    - 'how likely is model to guess the sentence correctly?'
    - simplifying assumption that reader's goal is accurate identification

Simulation 1: Methods

Task

- simulate reading on sentences from the Schilling et al. (1998) corpus
- measure average timesteps T and accuracy L

Simulation 1: Methods

Task

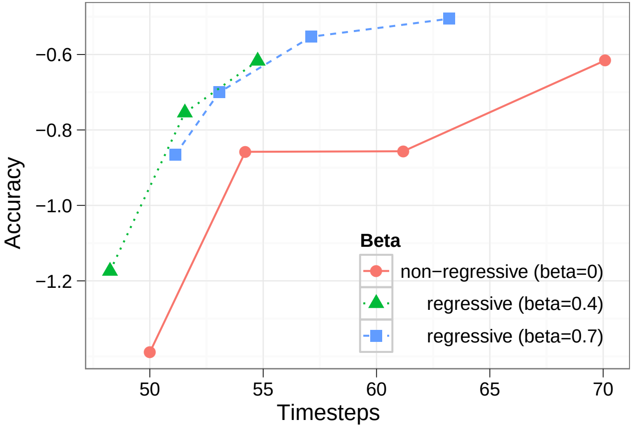
- simulate reading on sentences from the Schilling et al. (1998) corpus
- measure average timesteps T and accuracy L

Model

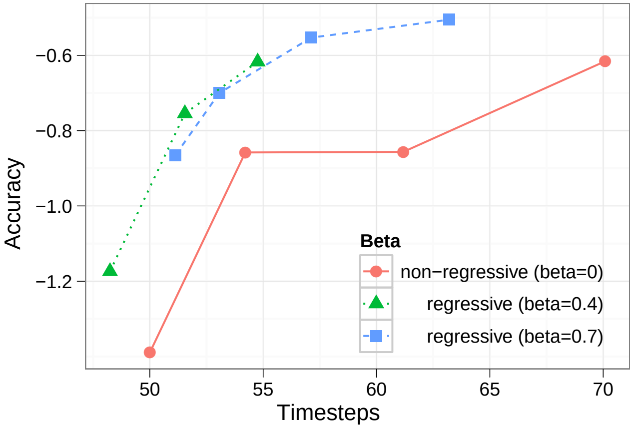
- bigram language model trained on British National Corpus
- implemented with weighted finite state automata using OpenFST (Allauzen et al., 2007)

Simulation 1: Results

Simulation 1: Results



Simulation 1: Results



Hypothesis confirmed: for each non-regressive policy ($\beta = 0$), there is a regressive policy that is both more accurate and faster

Simulation 2

Motivation

- one advantage of this framework: actions are goal-directed
- thus, we can ask a new question:
  - How should reading behavior change depending on the reader's goal?

Simulation 2

Motivation

- one advantage of this framework: actions are goal-directed
- thus, we can ask a new question:
  - How should reading behavior change depending on the reader's goal?

Specifically, using simple policy family and simple goal functions

- simple policy family: how should α and β change?
- goal functions: relative values of timesteps T vs. accuracy L

Simulation 2

Motivation

- one advantage of this framework: actions are goal-directed
- thus, we can ask a new question:
  - How should reading behavior change depending on the reader's goal?

Specifically, using simple policy family and simple goal functions

- simple policy family: how should α and β change?
- goal functions: relative values of timesteps T vs. accuracy L

Simulation 2: find optimal values of α and β for three goal functions

Simulation 2: Methods

Formal goal functions

- consider linear combinations of L and T: $L(1-\gamma) - T\gamma$
  where $\gamma \in [0,1]$ gives weighting of time
- we optimize for 3 goals: $\gamma \in \{.025, .1, .4\}$

Simulation 2: Methods

Formal goal functions

- consider linear combinations of L and T: $L(1-\gamma)-T\gamma$ where $\gamma \in [0,1]$ gives weighting of time
- we optimize for 3 goals: $\gamma \in \{.025, .1, .4\}$

Optimization

- use the PEGASUS method (Ng & Jordan, 2000) to estimate performance for a given $[\alpha, \beta]$
- (essentially, just average performance over sentences from Schilling)
- then, we use standard hillclimbing techniques to find optimum

Simulation 2: Methods

Formal goal functions

- consider linear combinations of L and T: $L(1-\gamma) - T\gamma$
  where $\gamma \in [0,1]$ gives weighting of time
- we optimize for 3 goals: $\gamma \in \{.025, .1, .4\}$

Optimization

- use the PEGASUS method (Ng & Jordan, 2000) to estimate performance for a given $[\alpha, \beta]$
- (essentially, just average performance over sentences from Schilling)
- then, we use standard hillclimbing techniques to find optimum

Model

- same as in Sim. 1, except trim the grammar to speed performance

Simulation 2: Results

γ	α	β
.025	.90	.99
.1	.36	.80
.4	.18	.38

Optimization worked

- α and β decrease as time valued more

Simulation 2: Results

γ	α	β	Timesteps	Accuracy (Prob.)
.025	.90	.99	41.2	-0.02 ($p \approx .98$)
.1	.36	.80	25.8	-0.90 ($p \approx .41$)
.4	.18	.38	16.4	-4.59 ($p \approx .01$)

Optimization worked

- α and β decrease as time valued more
- mean T and L decrease as time valued more

Simulation 2: Results

γ	α	β	Timesteps	Accuracy (Prob.)
.025	.90	.99	41.2	-0.02 ($p \approx .98$)
.1	.36	.80	25.8	-0.90 ($p \approx .41$)
.4	.18	.38	16.4	-4.59 ($p \approx .01$)

Optimization worked

- α and β decrease as time valued more
- mean T and L decrease as time valued more

Sidenote: confirmation of Sim. 1

- Optimal policies have positive β

Conclusion

A new framework for eye movements in reading

- readers achieve diverse goals by getting info about text identity
- text identity is normatively given by Bayesian inference combining
  - language knowledge (prior)
  - visual input (likelihood)

Conclusion

A new framework for eye movements in reading

- readers achieve diverse goals by getting info about text identity
- text identity is normatively given by Bayesian inference combining
  - language knowledge (prior)
  - visual input (likelihood)
- eye movements are produced to get visual input

Conclusion

A new framework for eye movements in reading

- readers achieve diverse goals by getting info about text identity
- text identity is normatively given by Bayesian inference combining
  - language knowledge (prior)
  - visual input (likelihood)
- eye movements are produced to get visual input

A new reason for regressions: to get visual input

Conclusion

A new framework for eye movements in reading

- readers achieve diverse goals by getting info about text identity
- text identity is normatively given by Bayesian inference combining
  - language knowledge (prior)
  - visual input (likelihood)
- eye movements are produced to get visual input

A new reason for regressions: to get visual input

- confidence will sometimes fall about previous regions
- Sim. 1: regressions are a rational solution to this problem
- Sim. 2: policies found by optimization use regressions

Future directions

Technical directions

- so far, very simple two-parameter behavioral policies
- investigating richer, more realistic classes of policies
- investigating using richer language models

Future directions

Technical directions

- so far, very simple two-parameter behavioral policies
- investigating richer, more realistic classes of policies
- investigating using richer language models

Modeling reader goals

- Sim. 2 demonstrated flexibility of model to adapt to different reader goals
- allows asking new questions:

Future directions

Technical directions

- so far, very simple two-parameter behavioral policies
- investigating richer, more realistic classes of policies
- investigating using richer language models

Modeling reader goals

- Sim. 2 demonstrated flexibility of model to adapt to different reader goals
- allows asking new questions: How should reading behavior change as
  - . . . accuracy is valued more or less relative to time?

Future directions

Technical directions

- so far, very simple two-parameter behavioral policies
- investigating richer, more realistic classes of policies
- investigating using richer language models

Modeling reader goals

- Sim. 2 demonstrated flexibility of model to adapt to different reader goals
- allows asking new questions: How should reading behavior change as
  - ...accuracy is valued more or less relative to time?
  - ...readers learn comprehension questions all ask about direct object?

Future directions

Technical directions

- so far, very simple two-parameter behavioral policies
- investigating richer, more realistic classes of policies
- investigating using richer language models

Modeling reader goals

- Sim. 2 demonstrated flexibility of model to adapt to different reader goals
- allows asking new questions: How should reading behavior change as
  - ...accuracy is valued more or less relative to time?
  - ...readers learn comprehension questions all ask about direct object?
  - ...readers learn that the text uses a lot of difficult words?

Future directions

Technical directions

- so far, very simple two-parameter behavioral policies
- investigating richer, more realistic classes of policies
- investigating using richer language models

Modeling reader goals

- Sim. 2 demonstrated flexibility of model to adapt to different reader goals
- allows asking new questions: How should reading behavior change as
  - ... accuracy is valued more or less relative to time?
  - ... readers learn comprehension questions all ask about direct object?
  - ... readers learn that the text uses a lot of difficult words?
  - ... a number of other exciting questions!

Thanks!

Discussion & Feedback

- Jeff Elman, Tom Griffiths, Andy Kehler, Keith Rayner, Angela Yu
- Audience at this year's LSA
- Members of the Computational Psycholinguistics Lab

Funding

- NIH training grant T32-DC000041 from the Center for Research in Language at UC San Diego
- UC San Diego Academic Senate

Allauzen, C., Riley, M., Schalkwyk, J., Skut, W., & Mohri, M. (2007). OpenFst: A general and efficient weighted finite-state transducer library. In *Proceedings of the Ninth International Conference on Implementation and Application of Automata, (CIAA 2007)* (Vol. 4783, p. 11-23). Springer.

Connine, C. M., Blasko, D. G., & Hall, M. (1991). Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constraints. *Journal of Memory and Language*, *30*(2), 234–250.

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, *112*(4), 777–813.

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences*, *106*(50), 21086–21090.

Ng, A. Y., & Jordan, M. (2000). PEGASUS: A policy search method for large MDPs and POMDPs. In *Uncertainty in Artificial Intelligence*, *Proceedings of the Sixteenth Conference.*

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review*, *113*, 327–357.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, *26*, 445–526.

Schilling, H. E. H., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition*, *26*(6), 1270–1281.