Simulating Language
Lecture 1: introduction

Kenny Smith

kenny.smith@ed.ac.uk

What have computers got to do with linguistics?

- A lot of money has been put into building better interfaces to computers that use natural language:
    - speech synthesis, speech recognition, machine translation, ...
- We also use computers as tools for linguistics:
    - speech analysis (e.g. spectrograms), corpus analysis (e.g. counting words, discovering patterns), psycholinguistics (e.g. displaying stimuli), ...

Computers as platforms for *modelling*

- This course is *not* about engineering, or research toolkits
- Instead, we will be looking at **modelling**
- Key questions:

  - What is a model?
  - Why would we want to build models in linguistics?
  - Why would computers help?

Everything you need to know about the course:
www.lel.ed.ac.uk/lec/simlang2015

This is a **practical** course: lots of time in the lab, working with simulations

- You do not need to know how to program, but you do need not to be scared of computers, and willing to try things out

- We will be working in a simplified subset of **Python**

  We will supply the code for the practicals, but you will need to modify it to carry out the tasks on the worksheets.

  This isn't a programming course: we aren't going to teach you how to program, but we will teach you just enough to understand and use some simple models we provide. You will have to meet us half way: you'll get on much better if you get your hands dirty and try things out.

- Labs provide lots of opportunity for one-to-one help and discussion.

Lab classes: 2pm, 3pm or 4pm

- Labs take place in DSB 1.16
- 18 PCs
- We need to split up into two or three groups!
- Please mark your preferences on the website, but note that we really need to spread the class over the three sessions as evenly as possible. Some re-jigging might be possible once everyone has indicated an initial preference.
- You need to always go to the same slot irrespective of each day (otherwise it'll be just too confusing!)

Assessment

- All aspects of the course will be assessed (i.e. every lecture, and every lab)
- Two take-home “exams”. Short-answer questions.
- Schedule of assessment release, deadlines, feedback dates available on the course webpage
- For UG students, each worth 50%. For PG students, second worth more.

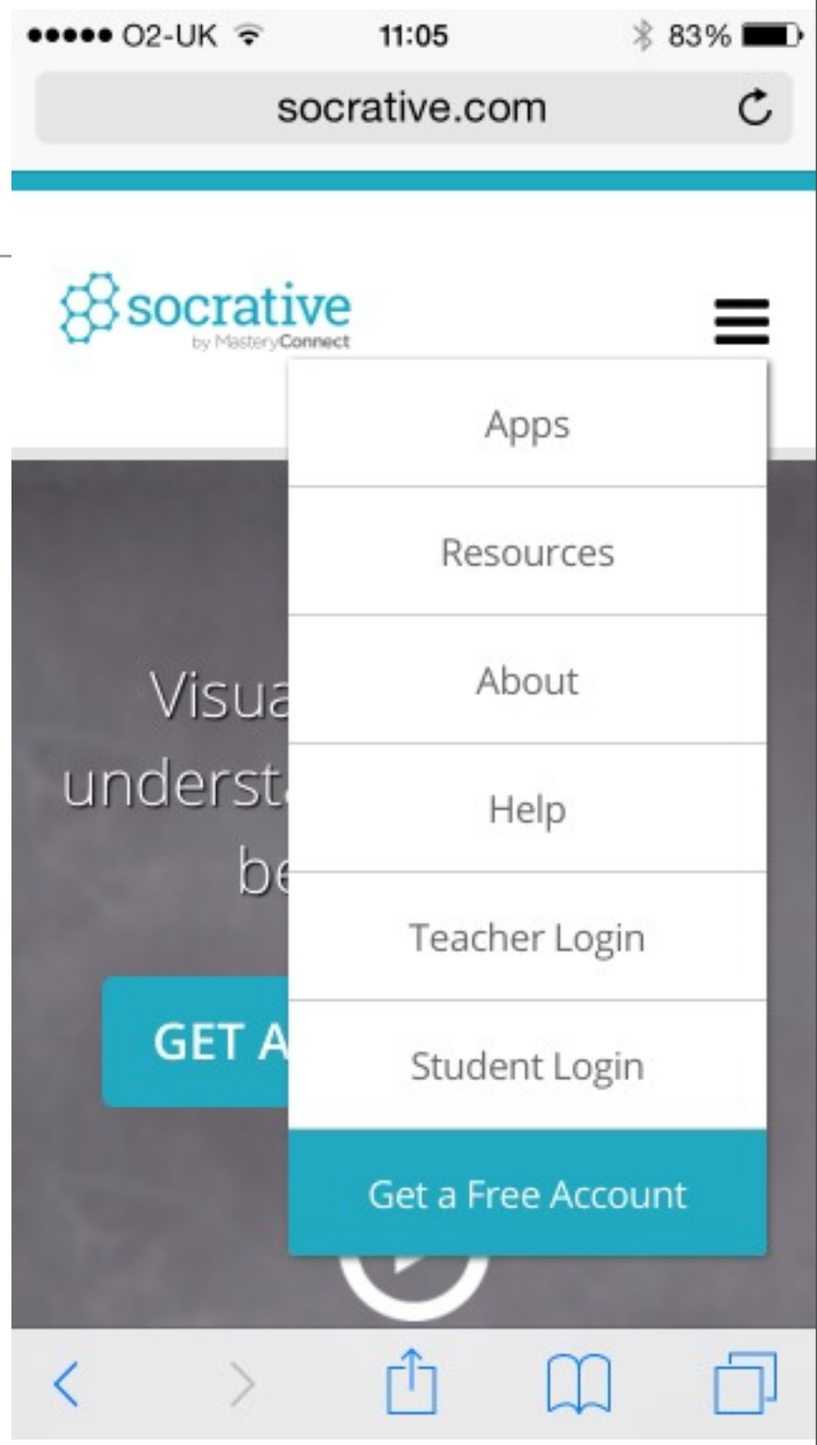
Readings

- Readings required, specified on the website
    - Do the readings before the lectures
    - Do the quiz (the most useful bit for me is the comment box at the end)

- Lab worksheets available online 1 week in advance
    - Attempt the worksheet before the lab, then attend the lab

Questions in lectures (hopefully)

- **Ask questions whenever you want to**
- But sometimes I'd like to ask you questions too
  - Take out your smartphone, laptop, whatever
  - Go to **socrative.com**
  - Log in as a student, enter room number **1f2864a3**



- **First question: I did the reading for today's lecture, true or false**

Alternative low-tech voting system!

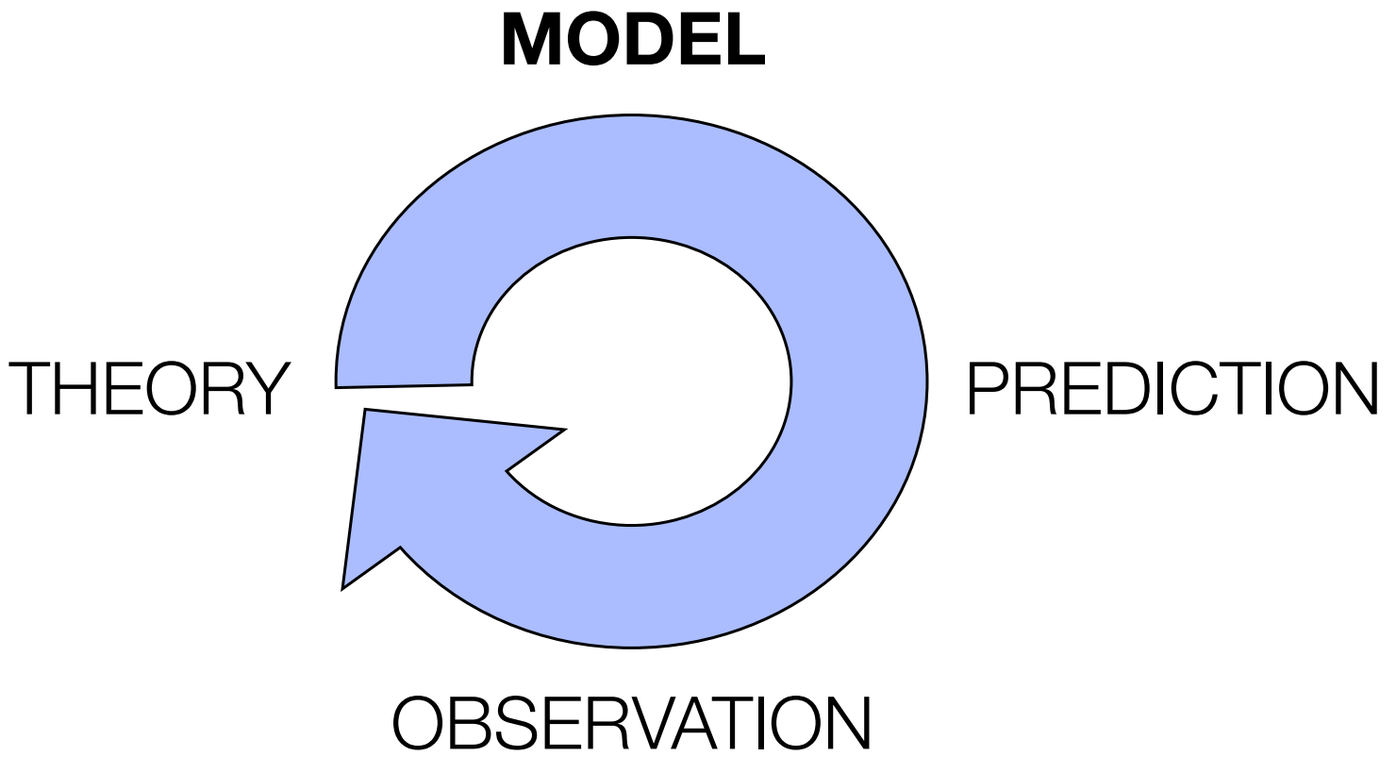
- Right hand across chest = True, left hand across chest = False
- ABCD = 1, 2, 3 or 4 fingers

What is a model?

- A miniature version of the system you are interested in
- Gains
  - Simplicity
  - Control
  - Ease of observation
- Loses: ?

What is a model for?

- One view:

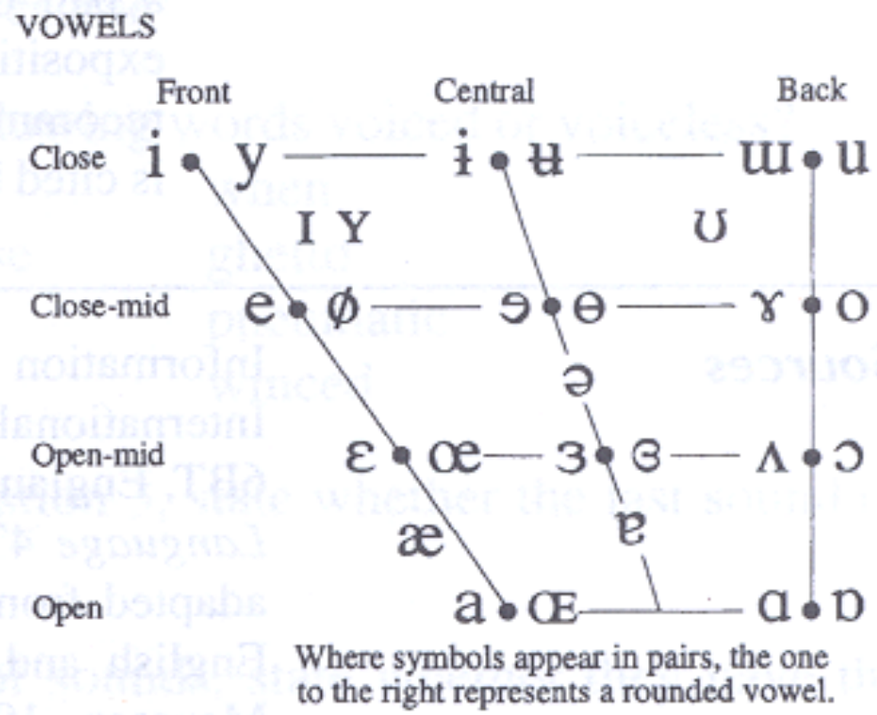


- We use models when we can't be sure what our theories predict

- Especially useful when dealing with *complex systems*

A simple example

- Vowels exist in a “space”



- Only some patterns arise cross linguistically.
  e.g. Vowel space seems to be symmetrically filled.

  Why might such universal patterns exist?

The need for a theory

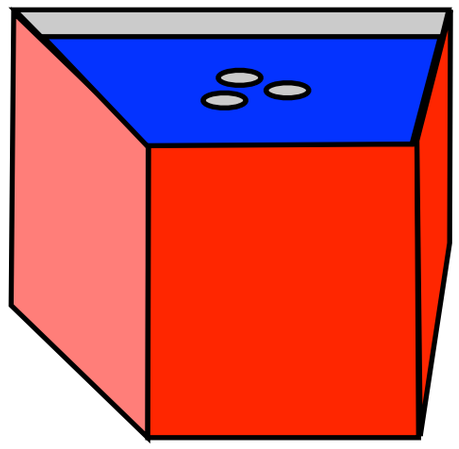
- Possible theory:
  - **Vowels tend to avoid being close to each other in order to maintain perceptual distinctiveness.**

- How do we tell if our theory is correct?

The need for a model

- We need some way of generating predictions from the theory which can be compared with the real data.

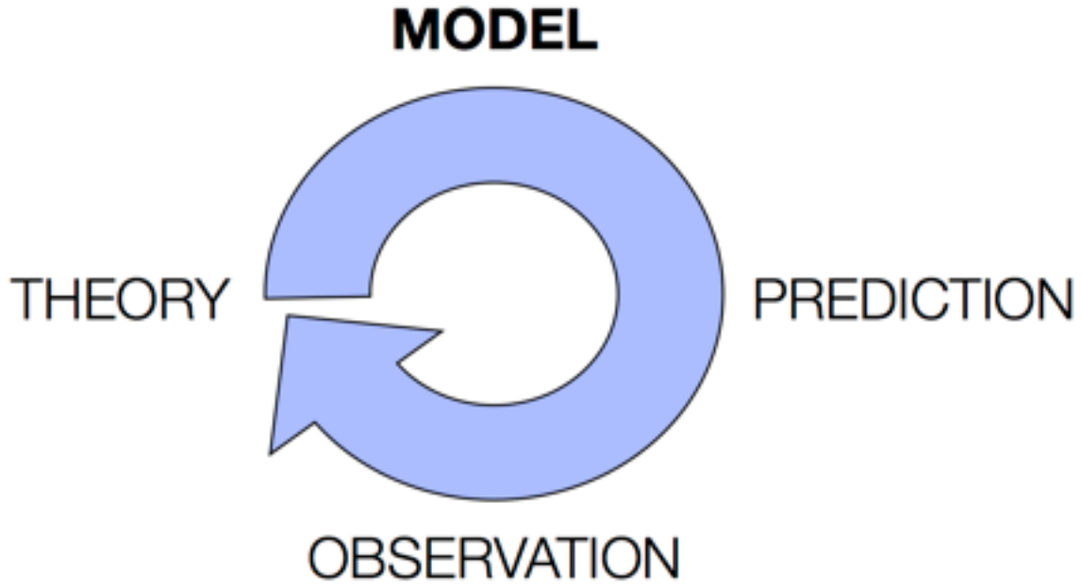
- A physical model (Liljencrants and Lindblom 1972): a tank of water, some corks with magnets attached.



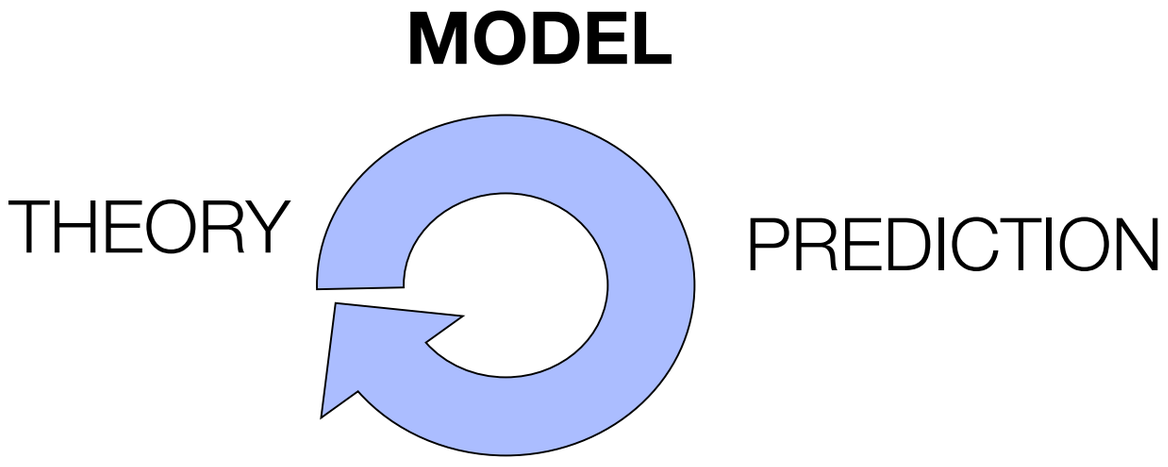
- From this model, predicted patterns can be compared with cross-linguistic data

I asked: have you seen models before?

- Most people said **no**
- But what about e.g.
  - syntactic theories
  - phonological theories
  - social networks and diffusion of sociolinguistic variants
  - ...
- All typically verbal, but still models (I think)



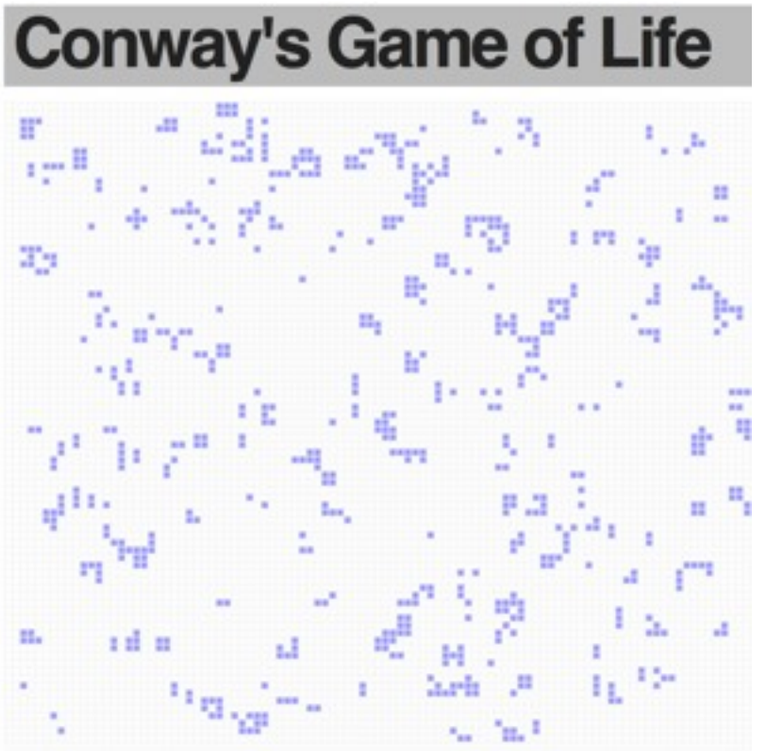
Revisited: What is a model for?



- An alternative / complementary approach: models as tools for understanding

  "Predictions are not the pinnacle of science. They are useful, especially for falsifying theories. However, predicting can't be a model's *only* purpose. ... surely the *insights* offered by a model are at least as important as its *predictions*: they help in understanding things by playing with them." (Sigmund, 1995, *Games of Life*, p. 4)

Life



- A ludicrously simple model (a cellular automaton)
  - Any cell with 3 live neighbours becomes live
  - Any live cell with 2 or 3 live neighbours stays alive
- Enormously complex behaviours
  - Including potential for self-replicating life-forms
- **I asked you: prediction testing or learning by playing?**

Question for you: How close to reality should models be?

- Reminder: socrative.com, room number 1f2864a3. You might need to refresh.

A: The model should leave nothing out, it should be as close as possible to reality

B: The model should leave lots of stuff out, it should be as simple as possible

My opinion

- What will we learn by building something as complex as the real thing?
- Simpler is better: easier to build, easier to run, better insights
- **Build your model to have as little extra in it that isn't part of your theory**
- We can come back to this

Why use computers for modelling?

- What if:
  - your theory is too difficult to understand simply through verbal argument, or introspection?
  - or a physical model cannot be constructed?
  - or a mathematical model is too difficult (or impossible) to construct?

- Particularly difficult problems involve **dynamic interactions**. For example:
  - a child's knowledge changing as she responds to hearing thousands of words
  - people interacting in groups over thousands of years
  - communicating organisms evolving over millennia

Computational modelling is the solution

- Computers are very good for models of many interacting components
- This has proved particularly valuable in allowing us to build a fundamentally *evolutionary* approach to understanding language
- In this course, we will be building and playing with models to tackle questions like:
  - How do innate signalling systems evolve?
  - How are communication systems shaped by cultural evolution?
  - Where do grammatical generalisations come from?
  - What do we mean when we say language is innate?

Next up

- Thursday: your first lab
    - Python basics: running python, variables, lists, conditionals, loops, functions, ...
- Friday: lecture on signalling
- Finally, my top tip for this course:

> Keep ahead with the readings, don't miss any lectures or labs! It'll be very tricky to keep up if you skip any sessions because everything builds on the previous step.