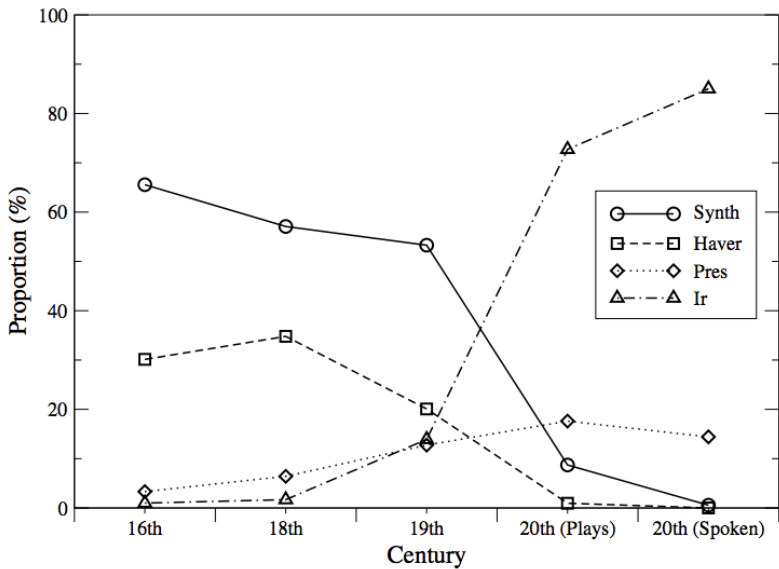


What's *that*?

Rob Truswell  
rob.truswell@ed.ac.uk

ELALT, Novi Sad, 9/3/19



# Change

- ▶ Language change tends to progress along an S-curve.
- ▶ We have a good story about why (Weinreich et al. 1968, Bailey 1973, Kroch 1989, Blythe & Croft 2012).
  - ▶ Forms compete to do the same job.
  - ▶ Something favours a particular form.
  - ▶ Gradually, that form spreads through the population.
- ▶ This story presupposes that the 'job' (or function) comes first, and that change involves finding a different way to do the same job (the **stable functions** assumption).
- ▶ If we remove that assumption, things change.

# Today

- ▶ An argument that the stable functions assumption isn't always warranted (partly joint work with Nik Gisborne).
- ▶ Plans for what to do without that assumption (joint work-in-progress with Richard Blythe, Simon Kirby).

## Two types of competition

- ▶ Part of language **use** is selecting among alternative forms which realize a communicative intention.
- ▶ Part of language **acquisition** involves pairing a given form with grammatical information.
- ▶ Both of these involve **competition**, but in different ways.
  - ▶ Among forms paired with a given function.
  - ▶ Among specifications of the function of a given form.

## The functions of functional heads

- ▶ Acquisition of content word meaning has been extensively investigated.
- ▶ But content words are the easy ones.
- ▶ Functional vocabulary is harder in many respects.
  - ▶ Ambiguity is the norm.
  - ▶ Mutual exclusivity not such a strong pressure.
  - ▶ Miscommunications less obvious and/or less serious.
  - ▶ Pairings between category and denotation more fluid.
- ▶ Learners are quick to figure out that *that* is a word.
- ▶ But it is much harder for them to answer a question like 'What is *that*?'.

## Worse

- ▶ The denotation of a lexical item doesn't directly determine what functions it can realize.
- ▶ That's determined by compositional interactions between lexical items (including an unspecified number of null lexical items), and by the many–many relationship between denotations and communicative intentions.
- ▶ Moreover, among functional vocabulary, polysemy is the norm, so figuring out *the* denotation of any relevant item is not easy.

## Change is change in associations

- ▶ Most well-studied cases of grammar change involve:
  - ▶ A stable set of forms
  - ▶ A stable set of functions
  - ▶ A dynamically changing set of alignments of forms with functions.
- ▶ *Do*-support emerged, but *gorp*-support never got off the ground: we rarely invent brand new grammatical lexemes to do extant jobs.
- ▶ And the set of jobs a grammar can do remains fairly stable (though not completely, e.g. Truswell & Gisborne 2016)
- ▶ Rather, grammars change because of novel answers to questions like 'What does *do* do?'.



## Innovations recur

- ▶ Innovative ‘mislearnings’ are not interesting or relevant, unless we can show that they’re not just noise.
- ▶ But the mislearnings aren’t random: they recur.

- (1) de fout **wie** hun eigenlijk maken  
the mistake who they actually make  
‘the mistake which they actually make’  
(Johan Cruyff, via Boef 2012)
- (2) adnominal adjectives (those **who** are not modifying the noun  
predicatively) (Belk 2016: 179)

## Grammar competition

- ▶ The problem with the stable functions assumption is that it doesn't allow for the full range of ways in which associations can change.
- ▶ In the general case, it's not immediately clear that competition-based explanations for phenomena related to S-curves are valid.

## Case study

- ▶ I'm going to talk about English relative clauses. Basic terms:
  1. Distinction between **headed** relatives (clauses modifying some external constituent, typically NP) and **free** relatives (clauses *with the function of* some other constituent, typically NP).
    - (3) a. I'll have the same thing [ $\emptyset$  that he's having \_\_\_]
    - b. I'll have [what  $\emptyset$  he's having \_\_\_]
  2. Both types of relative have dedicated **specifier** and **head** positions.
    - ▶ Possible specifiers: inflected demonstrative phrases in OE, *wh*-phrases,  $\emptyset$ .
    - ▶ Possible heads: OE *þe*, *that*, marginally *as*,  $\emptyset$ .

Each position can be filled or empty independently of the other in either type of relative, at least at some point in the last 1,000 years.

## Case study

- ▶ We will try to understand the brief period in 13th-century English when virtually every relative clause (headed or free) was introduced by complementizer *that* with an empty specifier (*peak that*).
  - ▶ This wasn't true in Old or Very Early Middle English.
  - ▶ It hasn't been true since Middle Middle English.
- ▶ Part of this is straightforward.
  - ▶ Demonstrative relatives disappeared as inflected demonstratives disappeared (slowly).
- ▶ Part of it can be understood in standard S-curve terms:
  - ▶ Between c.1150–1250, *þe*, which had been the most common complementizer in relative clauses, was replaced by *that*. *þe* and *that* are forms competing to realize the same function.
- ▶ Part of it (*wh*-relatives) only really makes sense when you consider competition among possible denotations of otherwise stable forms.

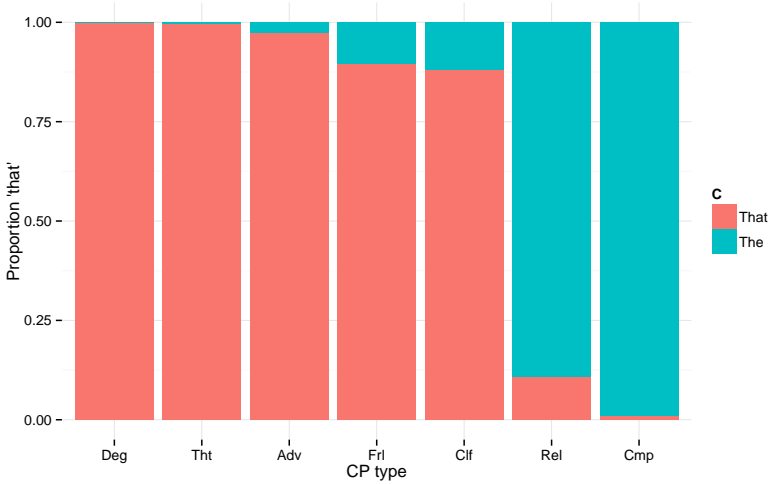
## Case study

- ▶ *Wh*-phrases in some contexts must be indefinite descriptions.
- ▶ In other contexts, they must be definite descriptions.
- ▶ The indefinite denotation is old, the definite denotation is newer.
- ▶ The change from indefinite to definite is possible because, within the scope of certain operators, it doesn't make much interpretive difference.
- ▶ (And the denotation of *wh*-forms in interrogatives may be neither of the above).
- ▶ The peak-*that* period corresponds to a lull between the death of indefinite *wh*-phrases and spread of definite *wh*-phrases.
- ▶ None of this can be explained by competition among forms (in some cases, *wh*-phrases aren't competing with anything) or by competition among functions (the non-isomorphism between denotations and functions is important).

## *þe* and *þæt*

- ▶ OE had two functionally specialized finite complementizers.
- ▶ *That* occurs in complement clauses, adverbial clauses (*if that*), degree clauses (*so much that*), most free relatives, most clefts.
- ▶ *þe* occurs in *the*-comparatives (*the more þe he ate*) and most headed relatives.
- ▶ This specialization is nearly categorical.

# $P_e$ and $\beta_{æt}$



## Relatives are messier

- ▶ The 'complementarity' is more of a strong tendency in relatives (including clefts).
- ▶ *Pe* sometimes occurs where *þæt* is expected. I don't know why.
- ▶ *þæt* sometimes occurs where *pe* is expected.
- ▶ In many cases, relativizer *þæt* is plausibly a demonstrative pronoun (*þæt*: DEM.N.SG.NOM/ACC).
- ▶ But in others, *þæt* displays the hallmark of OE relative complementizers (Allen 1977): P-stranding (Mitchell 1985).

(4) Pa ... næfde he scyld æt honda, þæt he þone  
When NEG.had he shield at hand that he the  
cyning [mid \_\_\_] scyldan meahte  
king with shield might  
'When . . . , he did not have a shield to hand with which  
he could shield the king.'

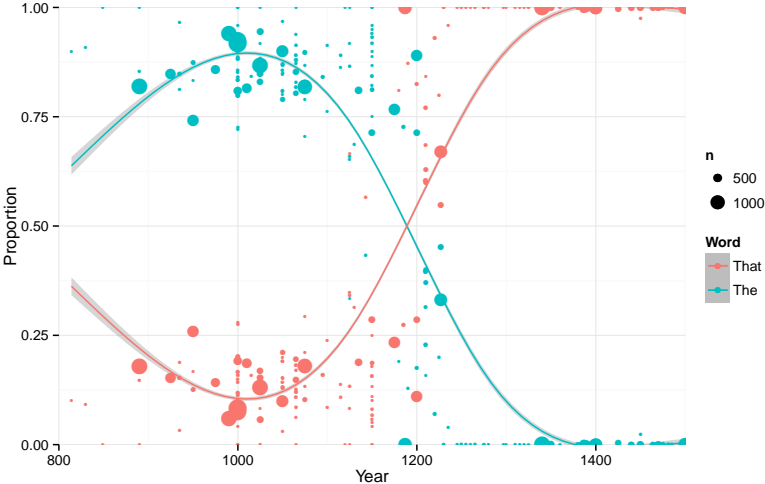
(cobede, Bede \_2:8.122.19.1160)



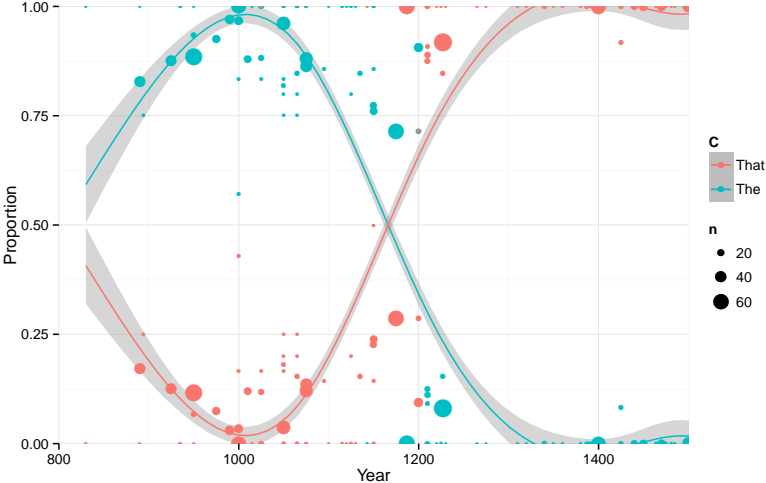
## Back-of-an-envelope calculation

- ▶ *pe* is undoubtedly a complementizer.
- ▶ It occurs in 16,846 OE relatives, of which 826 involve P-stranding (4.9%).
- ▶ There are 76 occurrences of relative *þæt* with P-stranding (out of 2,715 relative *þæt*).
- ▶ Although we can't know the incidence of relative-complementizer *þæt* as opposed to relative-specifier *þæt*, this suggests an estimate of  $76/0.049 = 1550$  occurrences in OE.
- ▶ This estimate suggests that *þæt* in relative clauses, even in OE, is normally a complementizer, not a demonstrative pronoun.
- ▶ So *pe* and *þæt* are in competition, and *þæt* wins.

# *Pe vs. $\beta_{æt}$ , all tokens*



# *Pe vs. $\beta_{æt}$ , P-stranding only*



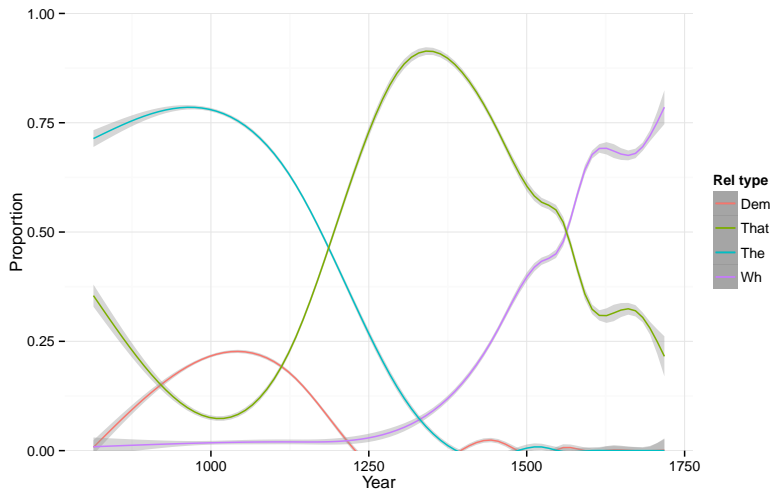
## *þe* vs. *þæt*: Summary

- ▶ *þæt* was always a complementizer.
- ▶ It was even always a relative complementizer, occurring as such with low frequency.
- ▶ Other demonstrative forms barely behaved like this (c.10x more common with *þæt* than other demonstratives).
- ▶ *þæt* killed *þe* within a couple of generations, following a trajectory that could be an S-curve (but so abrupt that the middle part of the trajectory is unclear).
- ▶ Compatible with classical grammar competition, though even here, distinctive transient grammars (see McIntosh 1948 on animacy effects in *Peterborough Chronicle*).

## A four-way fight

- ▶ The competition between *þe* and *þæt* overlapped with the loss of demonstrative relative specifiers, and subsequent introduction of interrogative relative specifiers.
- ▶ Neither the specifier nor the complementizer has to be present.
- ▶ So the bigger picture could be construed as:
  - ▶ a 4-way fight (DEM, *þe*, *þæt*, WH);
  - ▶ two simultaneous 3-way fights (DEM, WH,  $\emptyset \times \textit{þe}, \textit{þæt}, \emptyset$ );
  - ▶ a 9-way fight (crossing the two 3-way fights).
- ▶ For the sake of our sanity, we'll stick with a 4-way fight and hope we're not losing much.

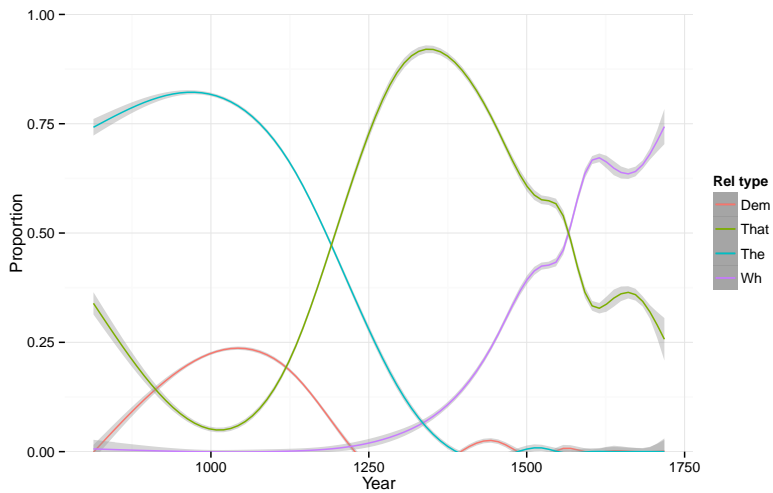
# The four-way fight over time



## Whatever

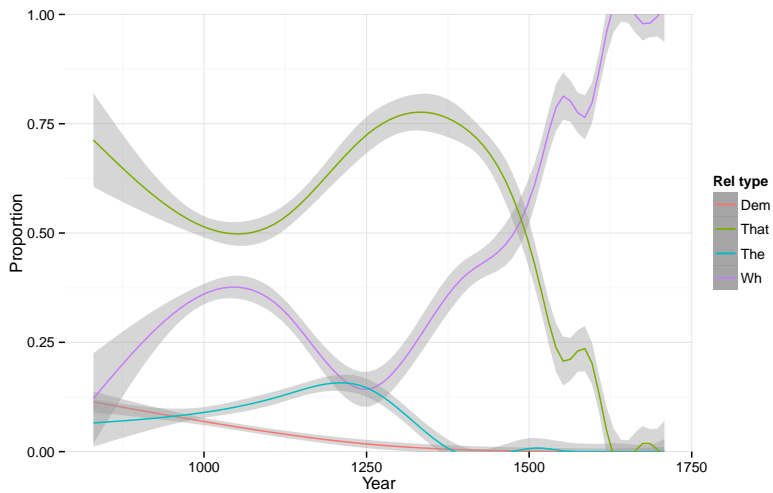
- ▶ This shows us the peak-*that* window: from c.1250–1400, around 90% of relatives were formed with *that*.
- ▶ This is much higher than before or since.
- ▶ But it doesn't look very interesting.
- ▶ *Pe* was the dominant strategy, then *that* became the dominant strategy, and it was even more dominant because the demonstratives (secondary strategy) died.
- ▶ However, a more interesting pattern is revealed when we factor out headed vs. free relatives.

# The four-way fight in headed relatives





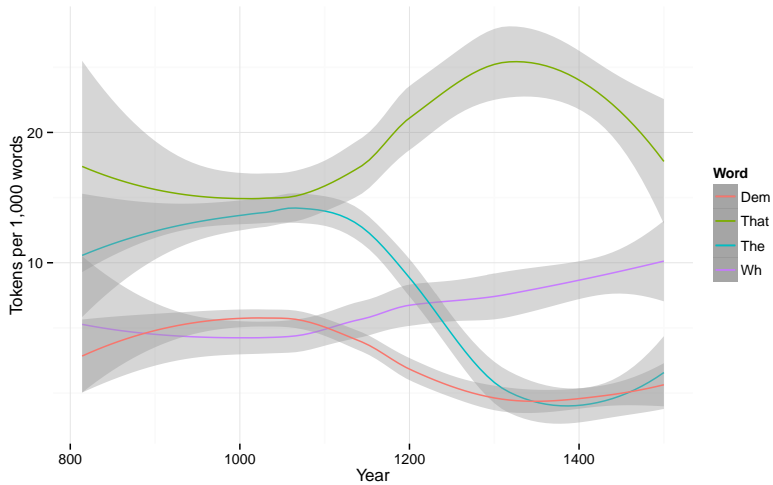
# The four-way fight in free relatives



## The N-shaped trajectory of free *wh*-relatives

- ▶ Free *hw*-relatives were well-established in OE.
- ▶ They slump appreciably in early ME.
- ▶ They rise again in late ME, and are now the only form of free relative.
- ▶ In its pomp, *that* not only killed *þe*, and replaced demonstratives in [Spec,CP], but took a chunk out of *wh* forms too.
- ▶ This is surprising, because *wh*-forms were in no danger of disappearing.

# Word frequencies over time



## Wh-form semantics

- ▶ We identify three different denotations for *wh*-forms.

### 1. Indefinite

- (5) and gif hwa hyt bletsað, þonne ablinð seo dydrung.  
and if who it blesses then ceases the illusion  
'and if anyone blesses it, then the illusion is dispelled.'  
(coaelhom,+AHom\_30:4.4082)

### 2. Definite

- (6) Gemyne, [hwæt Sanctus Paulus cwæð]  
remember what Saint Paul said  
'Remember what Saint Paul said.'  
(cogregdC,GDPref\_and\_3\_[C]:15.207.28.2739)

### 3. Interrogative

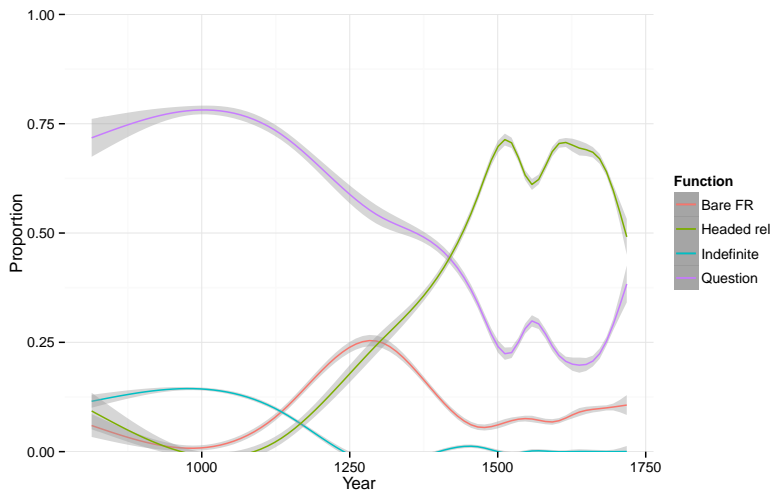
- (7) Hwær lede ge hine?  
where lead you him  
'Where are you leading him?' (coaelhom,+AHom\_6:77.915)

- ▶ We can track the diachronies of these denotations, by tracking the frequency of constructions which require one of them.

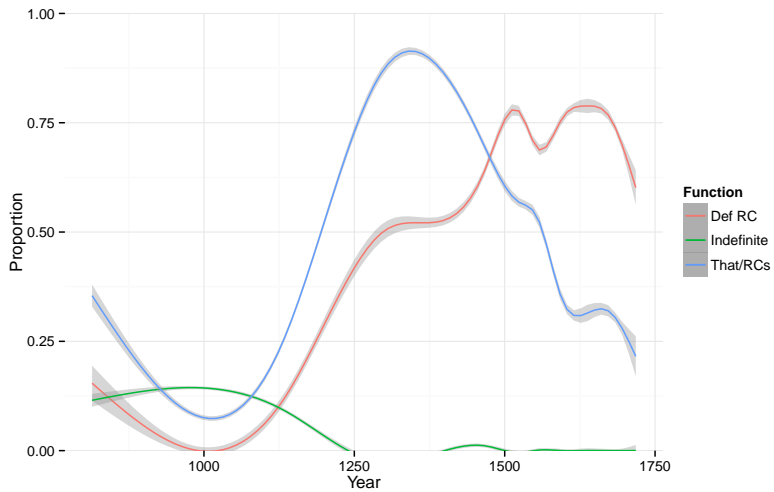
## Denotations and constructions

- ▶ *Wh*-indefinites are indicative of an indefinite denotation.
- ▶ Bare free relatives and nonrestrictive headed relatives are indicative of a definite denotation.
- ▶ In other constructions, the distinction is unclear or makes less sense (other free relatives, restrictive relatives, interrogatives).
- ▶ If it's less clear to us, it's less clear to the learner, and so ripe ground for change.
- ▶ Implication: relatively minor *wh*-constructions may be disproportionately important to a learner trying to figure out lexical meaning.

# Denotations over time



## Alignment with peak *that*



## Summary

- ▶ As we approach peak *that*, three things happen:
  1. Inflected demonstratives mainly disappear.
  2. *þe* mainly disappears.
  3. *Wh*-forms lose their indefinite denotation.
- ▶ The spread of definite *wh*-denotations to headed relatives brings an end to peak *that*.
  - ▶ Bare (definite) free *wh*-relatives increase in frequency in the 13th century. This clear trend doesn't much dent peak *that* because free relatives are relatively rare (headed relatives are 10x more common).
  - ▶ Headed *wh*-relatives follow 100 years later. Most early headed relatives are nonrestrictive.



## Implications for models of change

- ▶ A grammar is a set of **associations** between a set of expressions (forms) and a set of denotations.
- ▶ Grammar change is change in the set of associations.
- ▶ The sets of expressions and denotations themselves are often stable.
- ▶ Expressions with overlapping denotations, and multiple expressions per denotation, are both common (no clear mutual exclusivity pressure).
- ▶ The relationship between denotations and communicative functions is indirect — depends on what **else** the grammar generates.
- ▶ Speakers can **always** circumlocute — no functional vocabulary is strictly necessary.

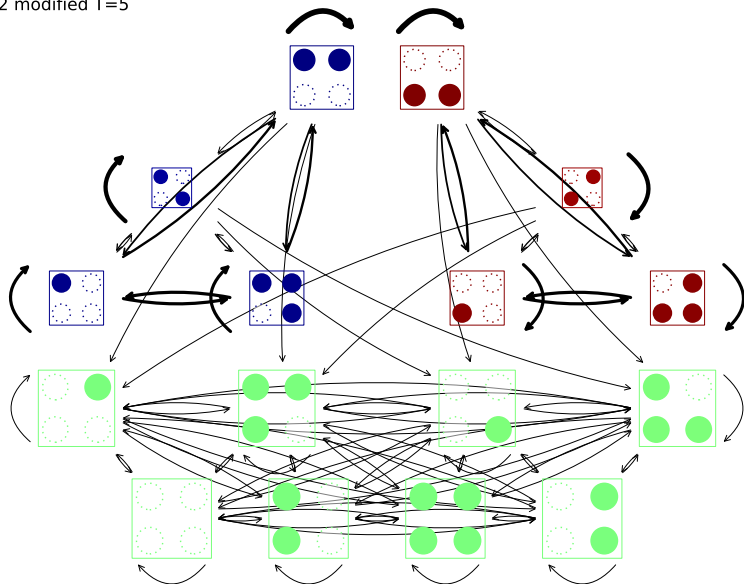
## The mutant dice model

- ▶  $f$  ‘meanings’ (or functions) and  $e$  expressions.
  - ▶ A meaning can be associated with 0– $e$  expressions.
  - ▶ An expression can be associated with 0– $f$  meanings.
- ▶ A grammar is a set of meaning–expression pairings.
- ▶ Flat prior over grammars.
- ▶ Meanings associated with different frequencies.
- ▶ There are  $T$  trials. In each trial:
  - ▶ An agent has to communicate about a given meaning.
  - ▶ The agent selects among expressions associated with that meaning (if there is one in the agent’s grammar), with small amount of noise.
  - ▶ The learner receives the form, and the intended expression with small amount of noise.
  - ▶ The learner updates the distribution over grammars accordingly.
- ▶ Two variants:
  1. If an agent doesn’t have an expression for a given function, pick an expression at random.
  2. An agent always has the option to circumlocute.

# Transition network

No circumlocution

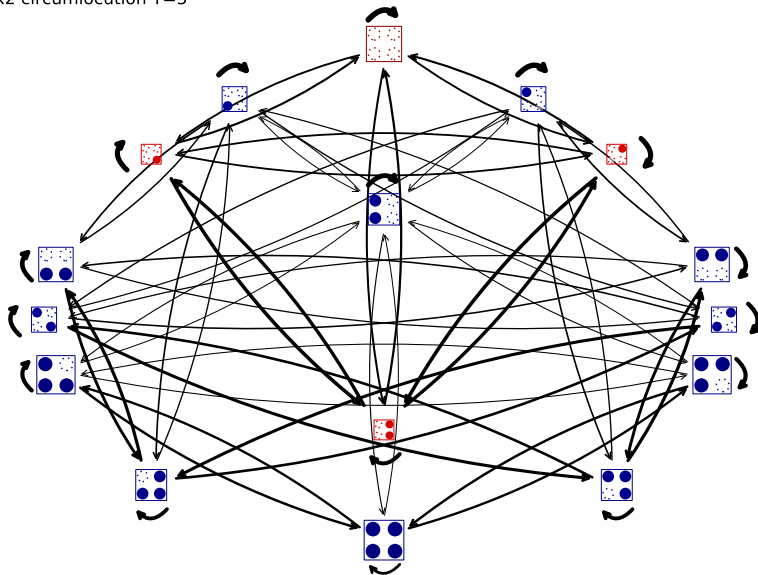
2x2 modified T=5



# Transition network

With circumlocution

2x2 circumlocution  $T=5$



## Visualization key

1. Each square represents a possible grammar, each arrow represents a change.
2. In each square, functions are columns and expressions are rows; functions ordered left-to-right by frequency.
3. lifetimes decrease from top to bottom.
4. equivalence classes on same row.
5. line thickness  $\propto$  transition probability
6. die size  $\propto$  stationary probability
7. colour  $\propto$  2nd eigenvector (blue = -ve, red = +ve, green  $\approx$  0).  
I've forgotten what this means.

## Pilot results

- ▶ In the circumlocution condition:
  - ▶ the most frequent transitions all involve changes in the expression of the less frequent function;
  - ▶ more frequent functions are more stably expressible.
- ▶ Neither applies to the model without circumlocution.
- ▶ No convergence to the prior: the prior is flat, the stationary distribution isn't (different dice are different sizes).
- ▶ Local lifetime is not correlated with stationary probability: no relation between thickness of arrow and bigness of dice.
- ▶ Most important part of all this: predictions about language typology that don't straightforwardly derive from the prior, or from the data, but rather from the dynamics of the system itself.
- ▶ Compare the three factors in Chomsky (2005) (innate stuff, experience, other).

## Next steps

- ▶ Try to refine these models to investigate links between:
  - ▶ Microscopic irreversibility (speaker has more knowledge of communicative intentions and ‘target grammar’ than learner);
  - ▶ Macroscopic irreversibility (nonrandom patterns of change, e.g. grammaticalization patterns, change relative to Accessibility Hierarchy).
- ▶ ‘Inverse problem’:
  - ▶ ‘develop statistical inferential methods to reverse-engineer the details of the individual grammar-learning process from trajectories of grammatical change in the historical record.’
  - ▶ Particularly the relative contributions of Chomsky’s three factors.
  - ▶ Might sound a bit ambitious, but at least this class of models has the right structure to allow dissociation of the three factors.

# Conclusions

- ▶ Grammar change is change in associations. The historical record says so.
- ▶ If you model grammar change in those terms, interesting emergent things start happening.
- ▶ Maybe one day we'll know what it all means.



# References

- Allen, C. (1977). *Topics in Diachronic English Syntax*. PhD thesis, University of Massachusetts, Amherst, MA.
- Bailey, C.-J. (1973). *Variation and Linguistic Theory*. Arlington, VA: Center for Applied Linguistics.
- Belk, Z. (2016). *Attributes of Attribution*. PhD thesis, University College London.
- Blythe, R. & Croft, W. (2012). S-curves and the mechanisms of propagation in language change. *Language*, 88, 269–304.
- Boef, E. (2012). *Doubling in Relative Clauses: Aspects of Morphosyntactic Microvariation in Dutch*. PhD thesis, Universiteit Utrecht.
- Chomsky, N. (2005). Three factors in language design. *Linguistic Inquiry*, 36, 1–22.
- Kroch, A. (1989). Reflexes of grammar in patterns of language change. *Language Variation and Change*, 1, 199–244.
- McIntosh, A. (1948). The relative pronouns *þe* and *þat* in Early Middle English. *English and Germanic Studies*, 1, 73–87.
- Mitchell, B. (1985). *Old English Syntax*. Oxford: Clarendon Press.
- Truswell, R. & Gisborne, N. (2016). A constant rate effect without stable functions. Paper presented at Evolang XI, New Orleans.
- Weinreich, U., Labov, W., & Herzog, M. (1968). Empirical foundations for a theory of language change. In W. Lehmann & Y. Malkiel (Eds.), *Directions for Historical Linguistics* (pp. 95–188). Austin, TX: University of Texas Press.