



Vrije  
Universiteit  
Brussel

# Weak-strong verb competition in Dutch

Dirk Pijpops, FCG-meeting August 2014



Vrije  
Universiteit  
Brussel



# The verb-inflection game

Modelling the Dutch strong-weak past tense inflection in  
an agent-based simulation

# Overview

---

- ▶ The strong-weak competition

  - How is it relevant?

- ▶ Basic design & alignment strategies

  - How does it work?

- ▶ Interpretation

  - What does it say?

- ▶ Conclusions & shortcomings

  - What can we conclude?

  - What should we improve?

---



# The strong-weak competition

---

- ▶ Strong ablaut alternation: e.g. *schuif* → *sch-oo-f*
- ▶ Weak suffix inflection: e.g. *schuif* → *schuif-de*
- ▶ Entrenchment vs. Transparency

⇒ Can these forces balance each other out, and if so, how?



# Basic Design: World

---

## ▶ 200 events

(  
  (0.105168148  gaan)  
  (0.18724294   moeten)  
  (0.267320342  zeggen)  
  ...  
  (0.908939552  trekken)  
  ...  
  (0.999941559  snuiten)  
  (0.999971429  wreken)  
  (1              schijten)  
)



# Basic Design: vi-agent

---

- ▶ Subclass of the Babel2 *agent* class
- ▶ 10 → 2 randomly selected to play a language game
- ▶ **Slots**
  - ▶ **Lexicon**
  - ▶ Name
  - ▶ String-utterance
  - ▶ Context
  - ▶ Topic
  - ▶ Produced-strong-verb-form-p
  - ▶ Produced-lexeme
  - ▶ Heard-lexeme
  - ▶ Total-weak-frequency
  - ▶ Delta
  - ▶ General-delta
  - ▶ Prestigious-p



# Basic Design: Lexicon

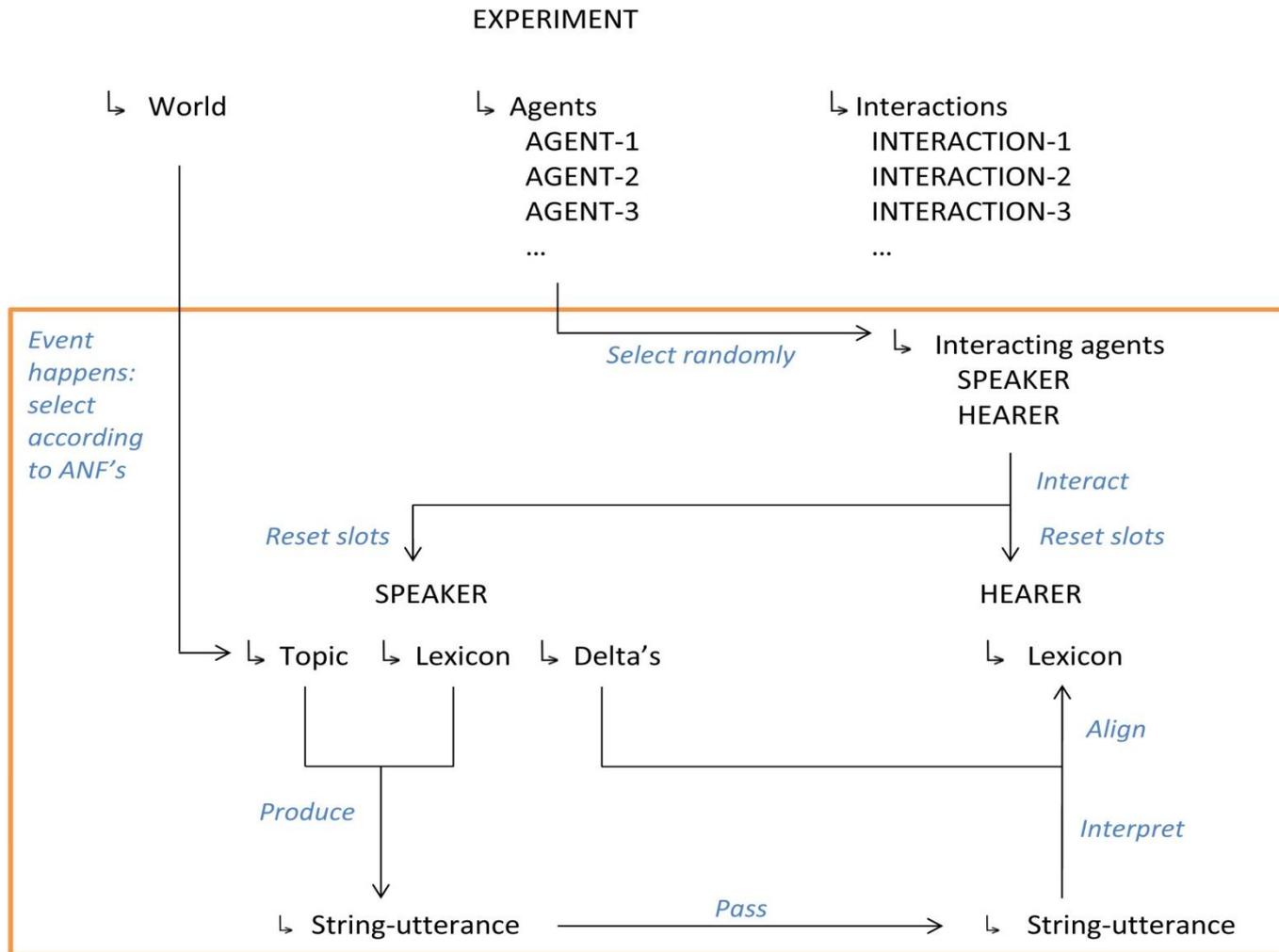
---

- ▶ Every agent
- ▶ Continuously updated

	<b>Event</b>	<b>Strong form</b>	<b>Strong frequency</b>	<b>Weak form</b>	<b>Weak frequency</b>	<b>Score</b>		
(	(	gaan	“ging”	0	“gade”	0	0.7	)
	(	moeten	“moest”	0	“moette”	0	0.7	)
	...							
	(	hangen	“hing”	0	“hangde”	0	0.7	)
	...							
	(	wreken	“wrook”	0	“wreekte”	0	0.7	)
	(	schijten	“scheet”	0	“schijtte”	0	0.7	)
)								



# Basic Design: Interaction



# Alignment Strategies

---

▶ How do agents update their lexicon after an interaction?

▶ Weak Transparency  $\frac{\delta_{gen}}{\delta_{spec}}$



# Alignment Strategies

---

## ▶ Frequency-based alignment

$$s_n = \frac{\text{Count}_{strong}(ev_n)}{\text{Count}_{strong}(ev_n) + \text{Count}_{weak}(ev_n) + \delta_{gen} \sum_{i=1}^{200} \text{Count}_{weak}(ev_i)}$$

## ▶ Delta multiplication alignment

	<b>Weak inflection</b>	<b>Strong inflection</b>
Verb-specific	$s_n = s_n (1 - \delta_{spec})$	$s_n = s_n (1 - \delta_{spec}) + \delta_{spec}$
General	$s_n = s_n (1 - \delta_{gen})$	

## ▶ Delta accumulation alignment

---



# Alignment Strategies: Delta accumulation

---

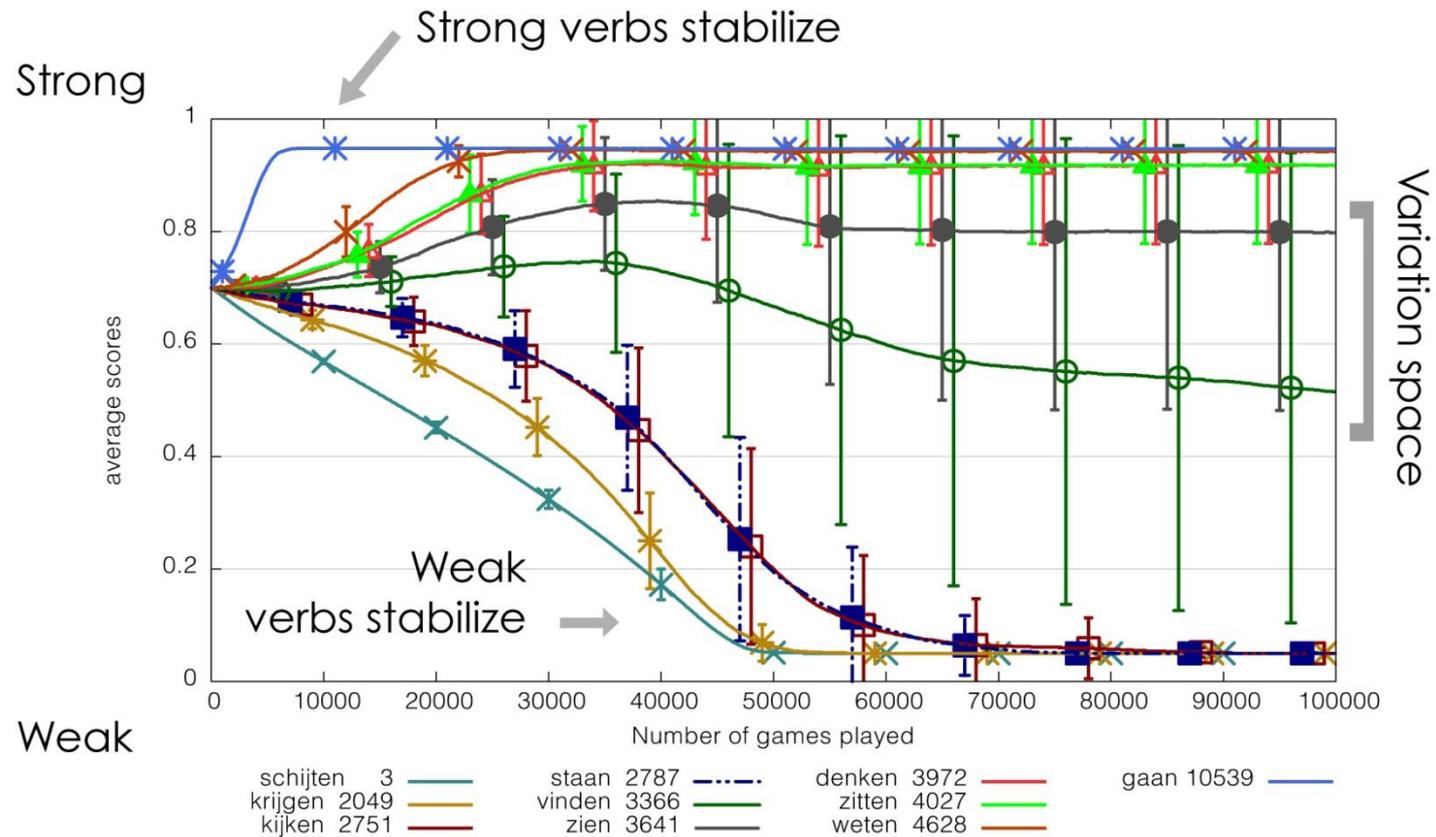
	<b>Weak inflection</b>	<b>Strong inflection</b>
Verb-specific	$S_n = S_n - \delta_{spec}$	$S_n = S_n + \delta_{spec}$
General	$S_n = S_n - \delta_{gen}$	

---



# Interpretation: Basic model

## ► Average score development

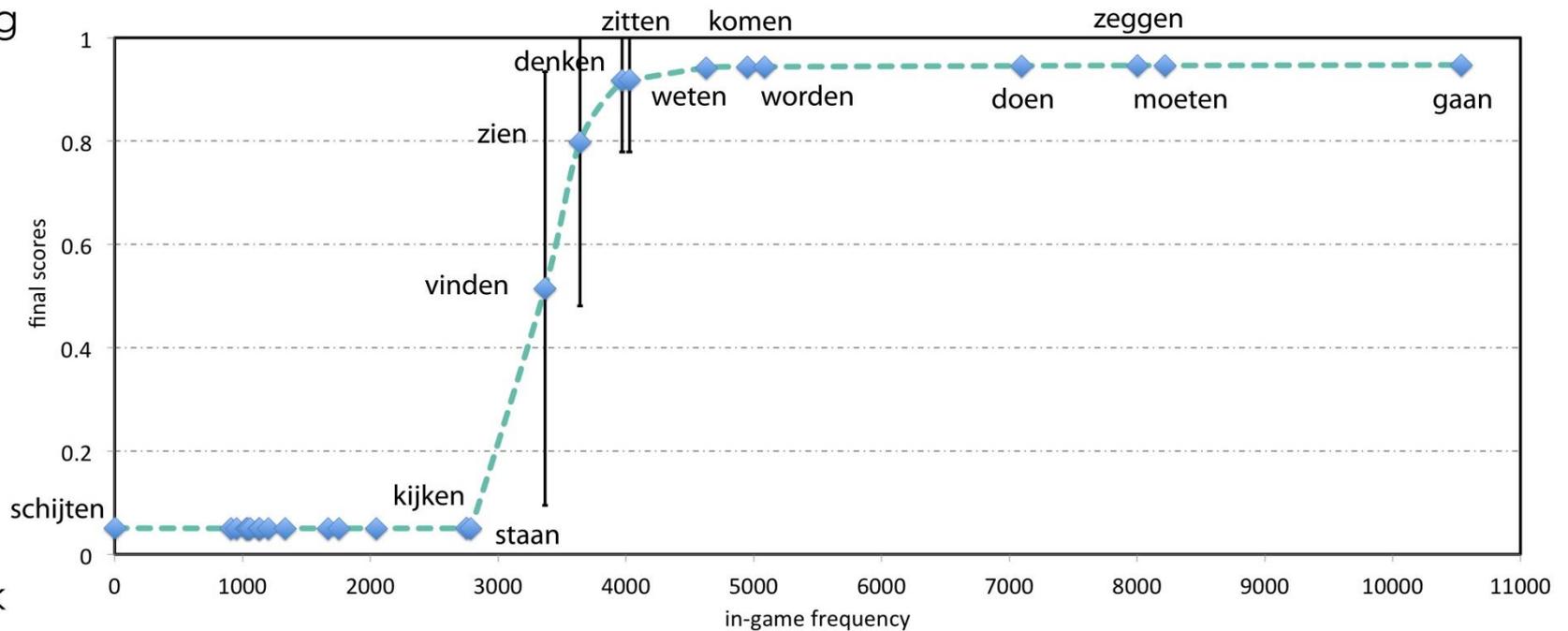


# Interpretation: Basic model

## ► Final scores

Variation space

Strong



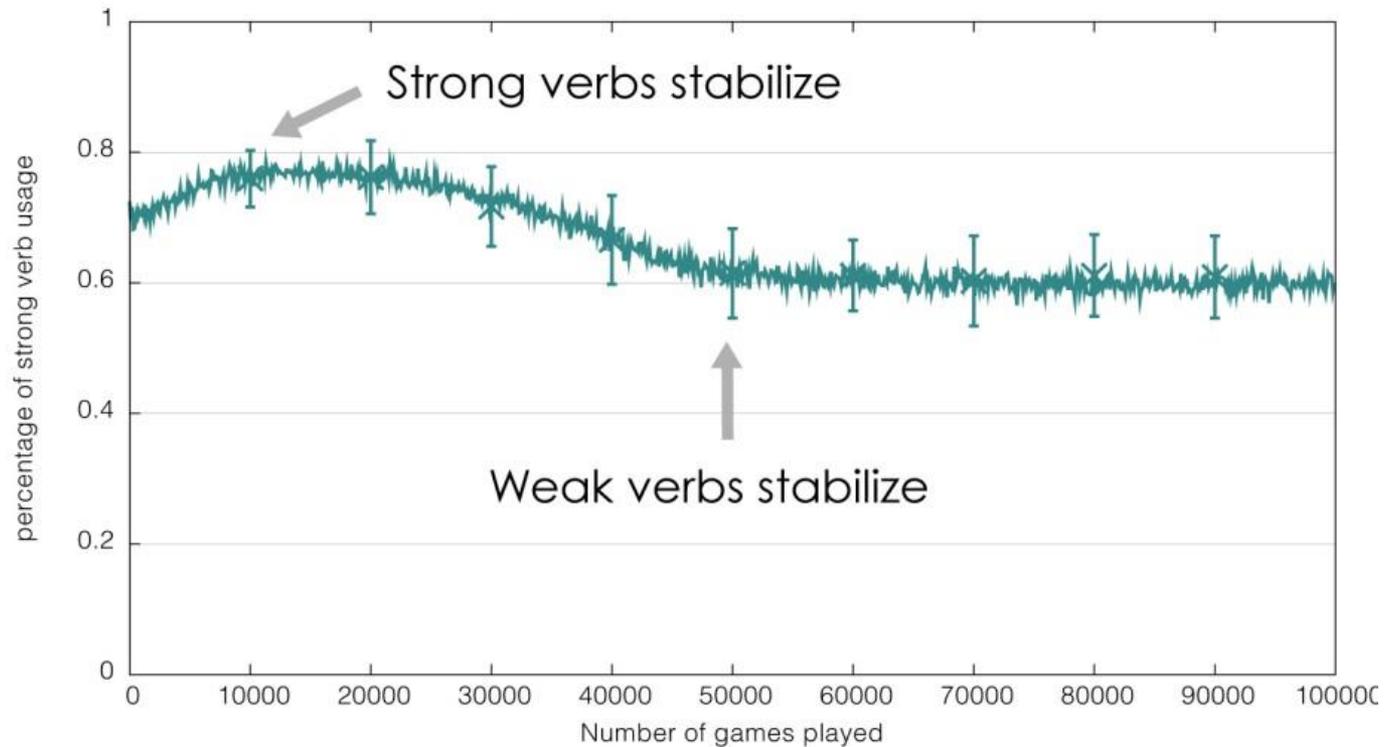
Weak



# Interpretation: Basic model

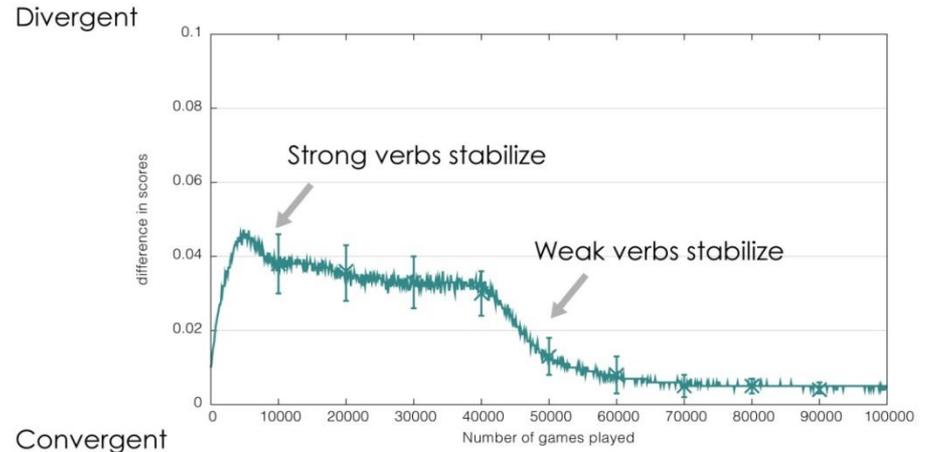
---

- ▶ Used inflection, i.e. token frequency

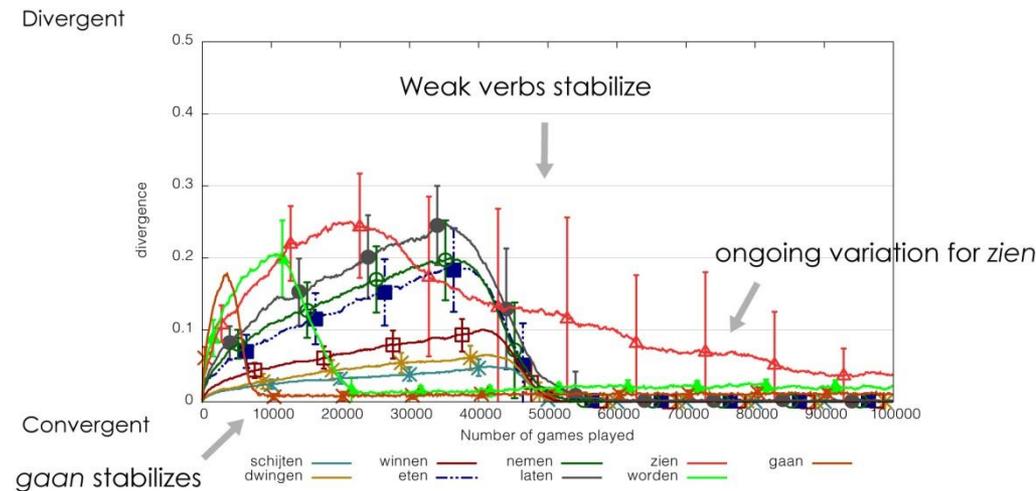


# Interpretation: Basic model

- ▶ Hearer-speaker convergence



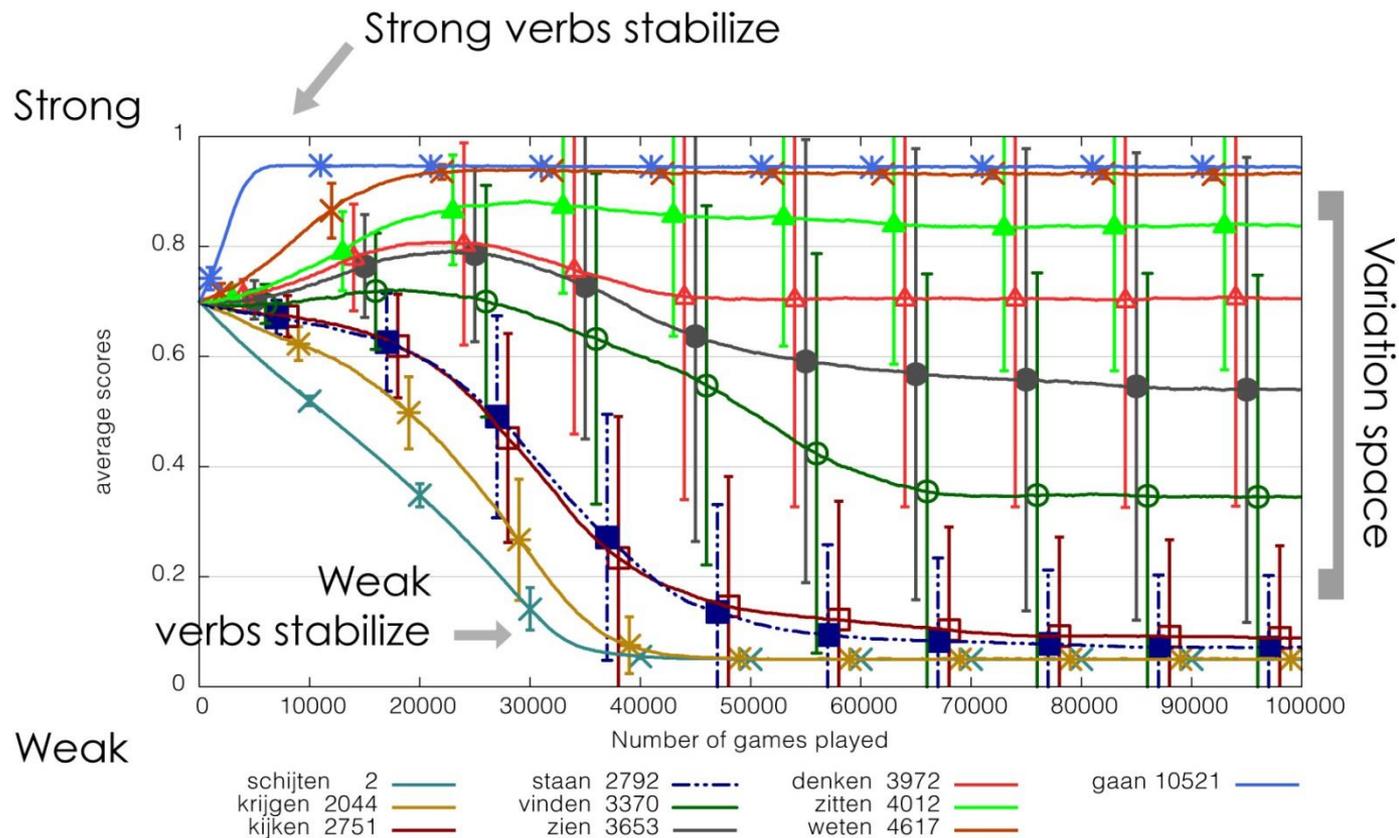
- ▶ Verb convergence





# Interpretation: Parameter changes

## ► Average score development



# Conclusions

---

- ▶ Goal internship accomplished
- ▶ Entrenchment vs. Transparency: delta accumulation
- ▶ Token frequency vs. Type frequency
- ▶ Prestigious agents
- ▶ Shortcomings: strong ablaut classes





Vrije  
Universiteit  
Brussel

Current work

# Strong ablaut classes

---

- ▶ Not completely untransparent & unproductive
- ▶ 7 historical classes (Strik & Knooihuizen)

Class	Pattern	Example	<i>n</i> types in ANS
I	/ɛI - e - e/	<i>grijpen - greep - gegrepen</i> ‘grab’	55 (53 <i>ij</i> + 2 <i>ei</i> )
IIa	/i - o - o/	<i>liegen - loog - gelogen</i> ‘lie (tell untruths)’	13
IIb	/œy - o - o/	<i>zuigen - zoog - gezogen</i> ‘suck’	24
IIIa	/I~ɛ - ɔ - ɔ/	<i>drinken - dronk - gedronken</i> ‘drink’	46 (26 <i>I</i> + 20 <i>ɛ</i> )
IIIb	/ɛ - i - ɔ/	<i>sterven - stierf - gestorven</i> ‘die’	7
IV	/e - a - o/	<i>spreken - sprak - gesproken</i> ‘speak’	6
V	/e - a - e/	<i>geven - gaf - gegeven</i> ‘give’	8
VI	/a - u - a/	<i>varen - voer - gevaren</i> ‘sail’	5
VIIa	X - /i/ - X	<i>blazen - blies - geblazen</i> ‘blow’	11
VIIb	/a - I - a/	<i>vangen - ving - gevangen</i> ‘catch’	2



# Goals

---

- ▶ **What makes a construction productive?**
  - ▶ Productivity of the strong verbs: Strik & Knooihuizen, Van de Velde
  - ▶ Type frequency effect
- ▶ **What is the future of the strong-weak competition?**
- ▶ **What are the effects of changing the parameters?**
- ▶ **Integrate corpus data, experimental results and agent-based modelling**



# Design: done

---

- ▶ Build a basic agent-based simulation
- ▶ Extract relevant frequencies & past tense forms from Corpus of Spoken Dutch
- ▶ Equip agents with FCG grammar containing a.o.
  - ▶ 1 weak inflection pattern
  - ▶ 12 strong inflection patterns
- ▶ Only finite forms



# Design: to do

---

- ▶ Build in competition
- ▶ Let the agents train on (part of) the language use of the corpus
- ▶ Adding new verbs



# Competition

---

- ▶ **Production**

- ▶ Apply constructions: lexeme, praeteritum, stem

- ▶ **Competition at the inflection level**

- ▶ Find all possible inflection patterns
  - ▶ Patterns enter in competition
  - ▶ Winner is used



# Competition

---

- ▶ **Known past tense**
  - ▶ Entrenchment of each pattern: token count
  - ▶ Personal preference of the verb: token count
- ▶ **Unknown past tense (productive use)**
  - ▶ Entrenchment of each pattern: token count
  - ▶ (Productivity of each pattern: type count)
- ▶ **Give probabilistic score**
- ▶ **Push winner to the top of the construction-supplier list**



# Plan of action

---

1. Program an agent-based simulation
2. Extract relevant frequencies from CGN
3. Design a working grammar
4. Build in a system of competition
5. Add monitors
6. Work away current restraints



# Problems

---

- ▶ Competition between applicable constructions
- ▶ Computational power when using realistic frequency ranges
- ▶ (Influence of completely irregular strong & weak verbs)



# Thanks to

---

- ▶ Dr. Katrien Beuls
- ▶ Dr. Pieter Wellens
- ▶ Dr. Freek Van de Velde
- ▶ All members of the AI-Lab
- ▶ Prof. Dr. Luc Steels
- ▶ Dr. Remi van Trijp
- ▶ You all



# References

---

- ▶ Beuls, K. & L. Steels. 2013. 'Agent-based Models of Strategies for the Emergence and Evolution of Grammatical Agreement'. *PLoS ONE* 8(3): e58960.
- ▶ Lieberman, E., Michel, J.-B., Jackson, B., Tang, T., & A. Martin. 2007. Quantifying the evolutionary dynamics of language. *Nature* 449 (7163): 713-716.
- ▶ Loetzsch, M., Wellens, P., De Beule, J., Bleys, J. & R. van Trijp. 2008. 'The Babel2 manual'. *AI-Memo 01-08*, AI-Lab VUB, Brussels, Belgium,
- ▶ Nübling, D. 2000. *Prinzipien der Irregularisierung: Eine kontrastive Analyse von zehn Verben in zehn germanischen Sprachen*. Tübingen: Max Niemeyer.
- ▶ Ringe, D. 2006. *From Proto-Indo-European to Proto-Germanic. A Linguistic History of English*. Volume I. Oxford: Oxford University Press.
- ▶ Strik, O. & R. Knooihuizen. In prep. 'Relative productivity potentials of Dutch verbal inflection patterns'.
- ▶ Van Eerten, L. 2007. 'Over het Corpus Gesproken Nederlands'. *Nederlandse Taalkunde* 12(3): 194-215.
- ▶ Van Trijp, R. 2013. 'Linguistic Selection Criteria for Explaining Language Change: A Case Study on Syncretism in German Definite Articles'. *Language Dynamics and Change* 3(1):105-132.
- ▶ Van Trijp, R. 2014. 'Fitness Landscapes in Cultural Language Evolution: A Case Study on German Definite Articles'. E.A. Cartmill, S. Roberts, H. Lyn & H. Cornish (eds.) *The Evolution of Language*. London: World Scientific. 334-342.
- ▶ Wellens, P. 2012. *Adaptive strategies in the emergence of lexical systems*. Brussels: Brussels University Press.

