

Strong islands of resilience in the weak flood.

Dirk Pijpops¹ & Katrien Beuls²

¹ Quantitative Lexicology and Variational Linguistics, University of Leuven

² Artificial Intelligence Lab, Vrije Universiteit Brussel

The Story

- Germanic past tense inflection (Grimm 1819)
 - Strong: wuif → w-oo-f
 - Weak: wuif → wuif-de

The Story

- Regularization ~ frequency of the verb
(‘weakenization’)

(Lieberman et al. 2007, Pugliesi et al. 2014, Pijpops et al. 2014)

- 2 additions
 - Strong classes
 - Sociohistorical context

(Carroll, Svare & Salmons 2012)

The Story

- ⇒ Not so hard: How can a pattern become dominant, if it is
 - the only regular one?
 - the most frequent one?

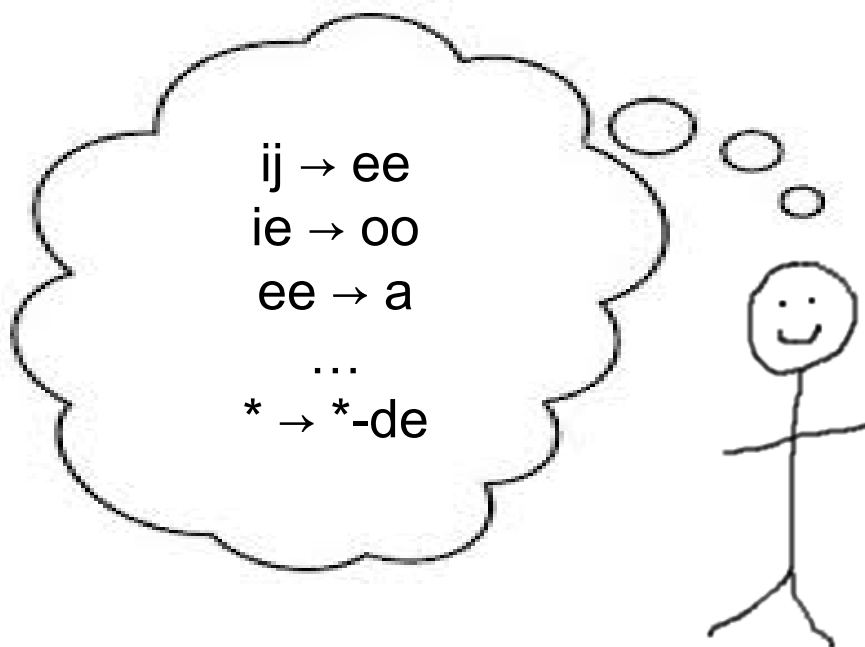
- ⇒ Harder: How can a pattern become dominant, if it is
 - neither the only regular one,
 - nor the most frequent one?

General Applicability

The Story

- Criticism

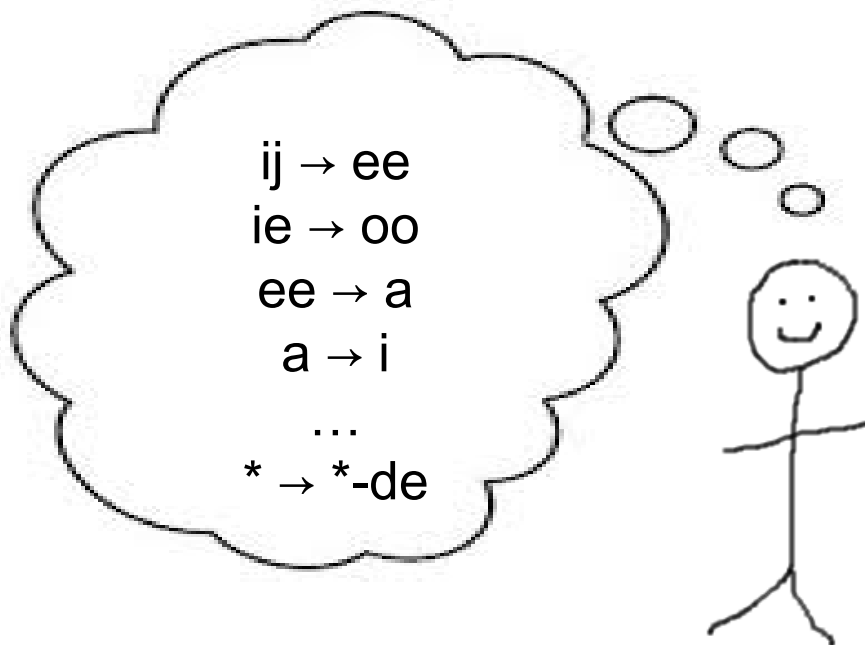
VANGEN



The Story

- Criticism

VANGEN



The Model

- Disclaimer: just testing the principle

The Model

- Agent-based
- 10 agents
- Past events
- FCG grammar: 11 strong patterns + 1 weak one
- Memory of previously heard forms
- Corpus Spoken Dutch

The Model

- Corpus Spoken Dutch

- Patterns

1	ij → ee	krijg	→	kreeg
2a	ie → oo	lieg	→	loog
2b	ui → oo	zuig	→	zoog
3a1	i → o	drink	→	dronk
3a2	e → o	zwem	→	zwom
3b	e → ie	sterf	→	stierf
4/5a	ee → a	spreek	→	sprak
5b	i → a	zit	→	zat
6	aa → oe	vaar	→	voer
7a	aa → ie	blaas	→	blies
7b	a → i	vang	→	ving
weak	* → *-te/de	hoor	→	hoorde

The Model

- Corpus Spoken Dutch
 - Patterns
 - 257 verbs
 - Starting Memory

The Model

- Competition formula:

SCHRIJVEN

$$p(\text{"schreef"}) \sim \text{freq}(\text{"schreef"}) + \text{freq}(\text{pattern 1})$$

$$p(\text{"schrijfde"}) \sim \text{freq}(\text{"schrijfde"}) + \text{freq}(\text{pattern weak})$$

The Model

- Competition formula:

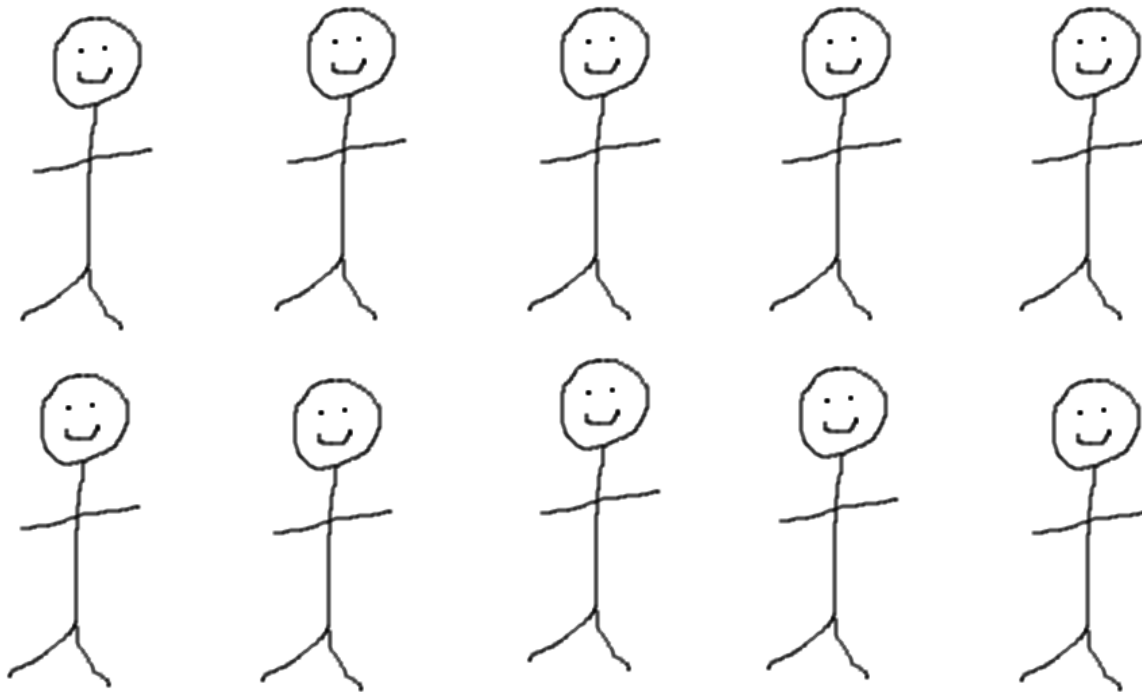
SCHRIJVEN

$$p(\text{"schreef"}) \sim \text{freq}(\text{"schreef"}) + \text{freq}(\text{pattern 1})$$

$$p(\text{"schrijfde"}) \sim \text{freq}(\text{"schrijfde"}) + \text{freq}(\text{pattern weak})$$

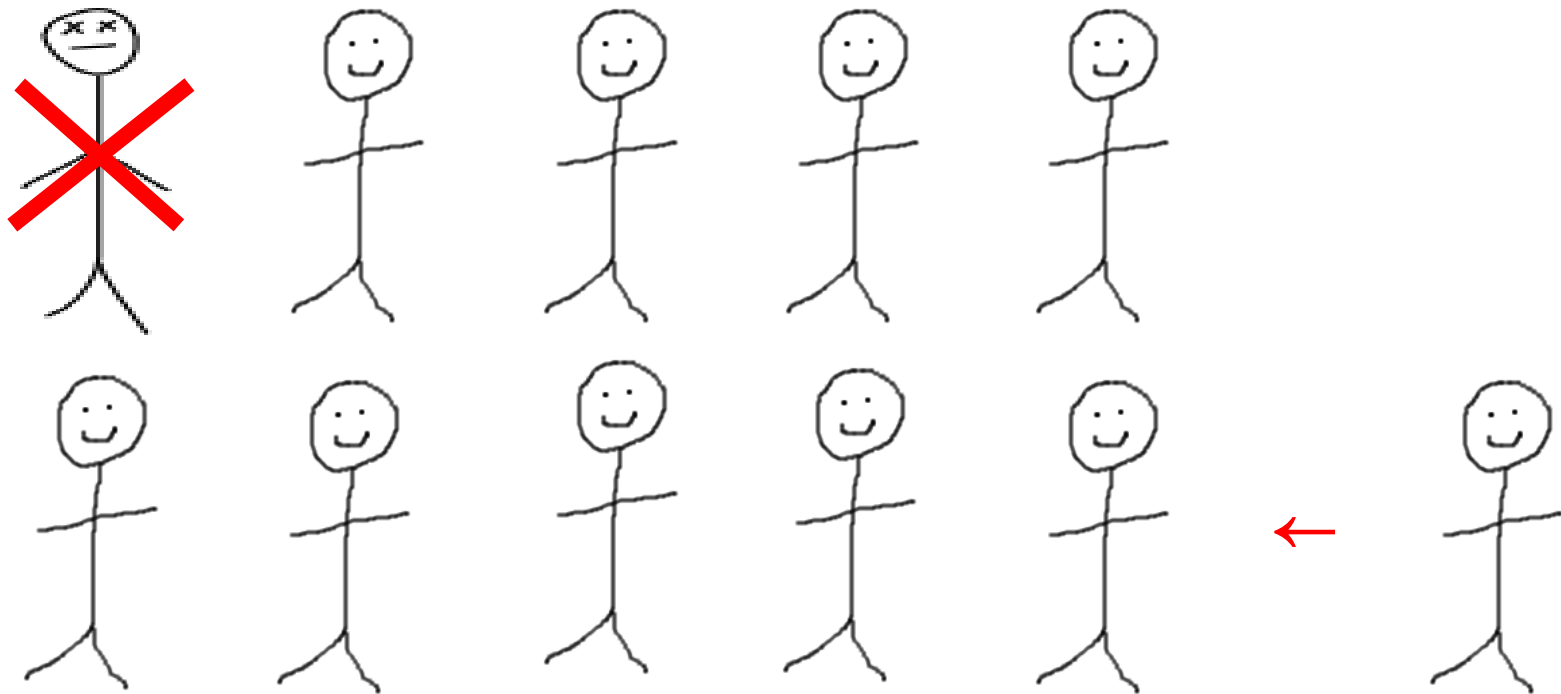
The Model

- Replenishment



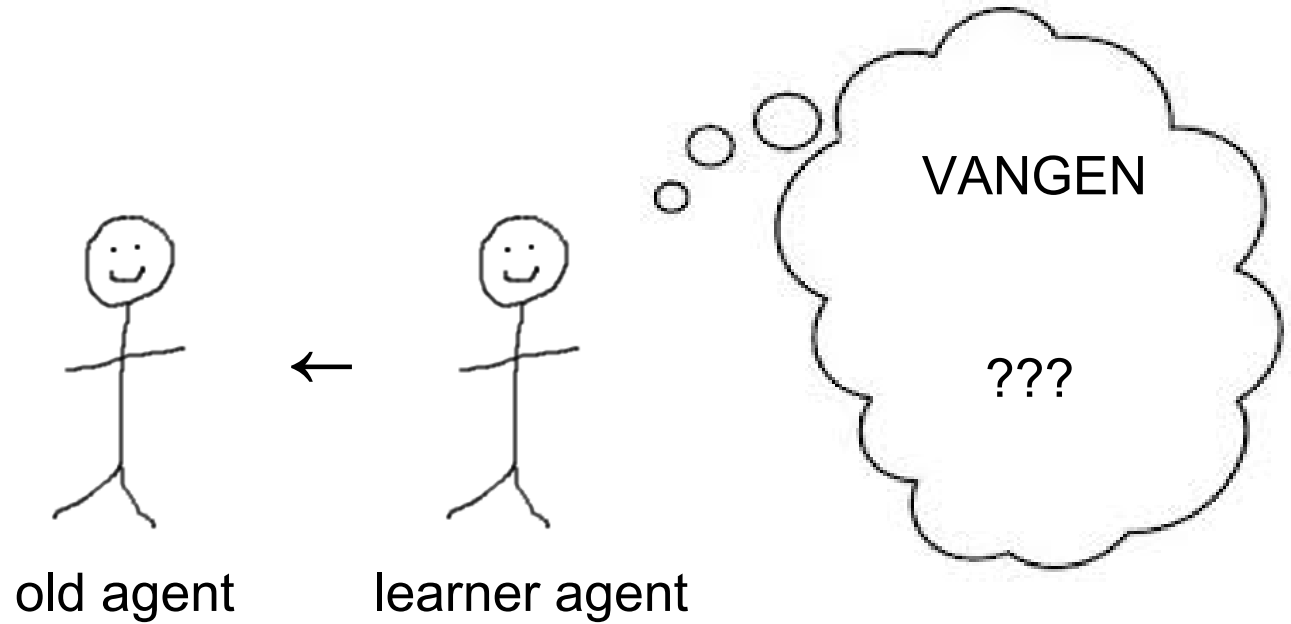
The Model

- Replenishment



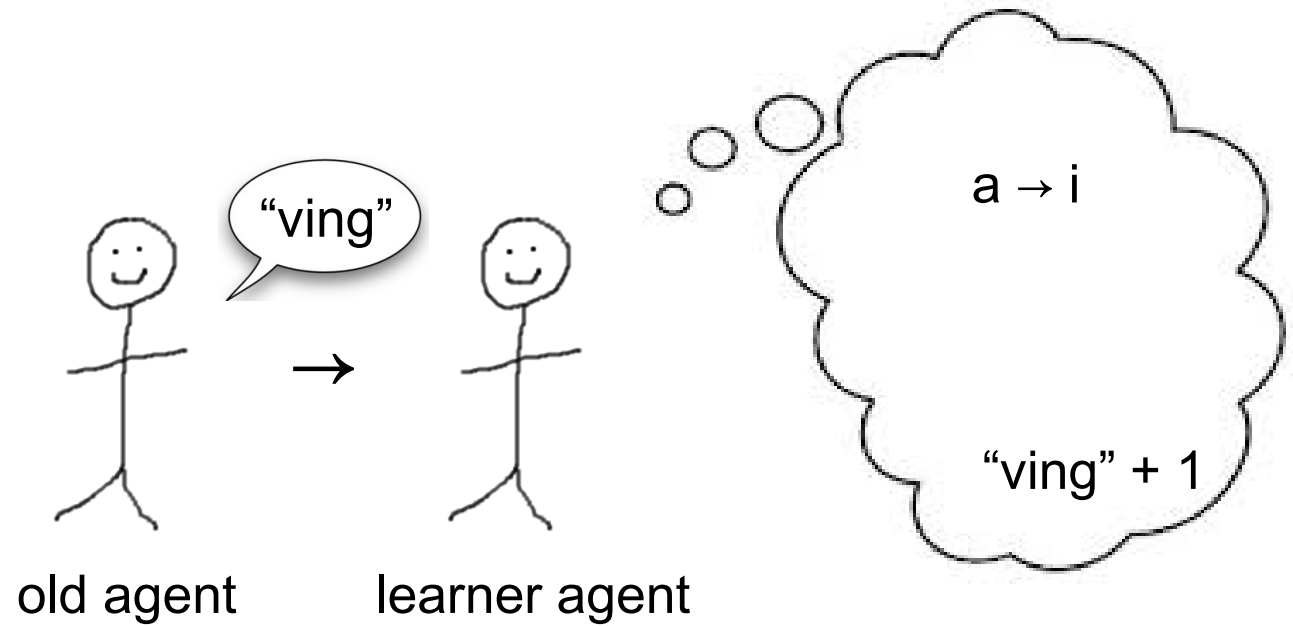
The Model

- Learning



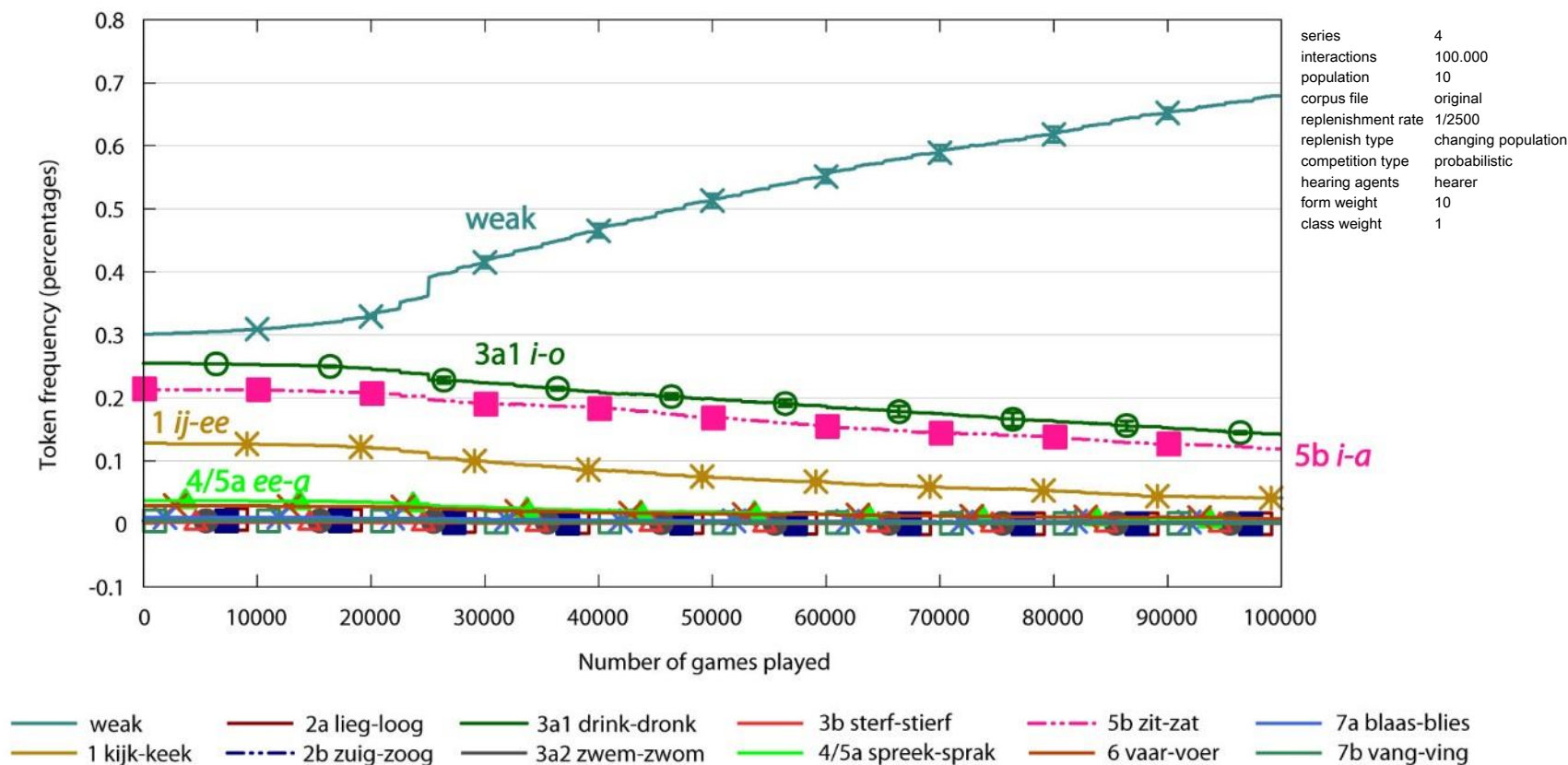
The Model

- Learning



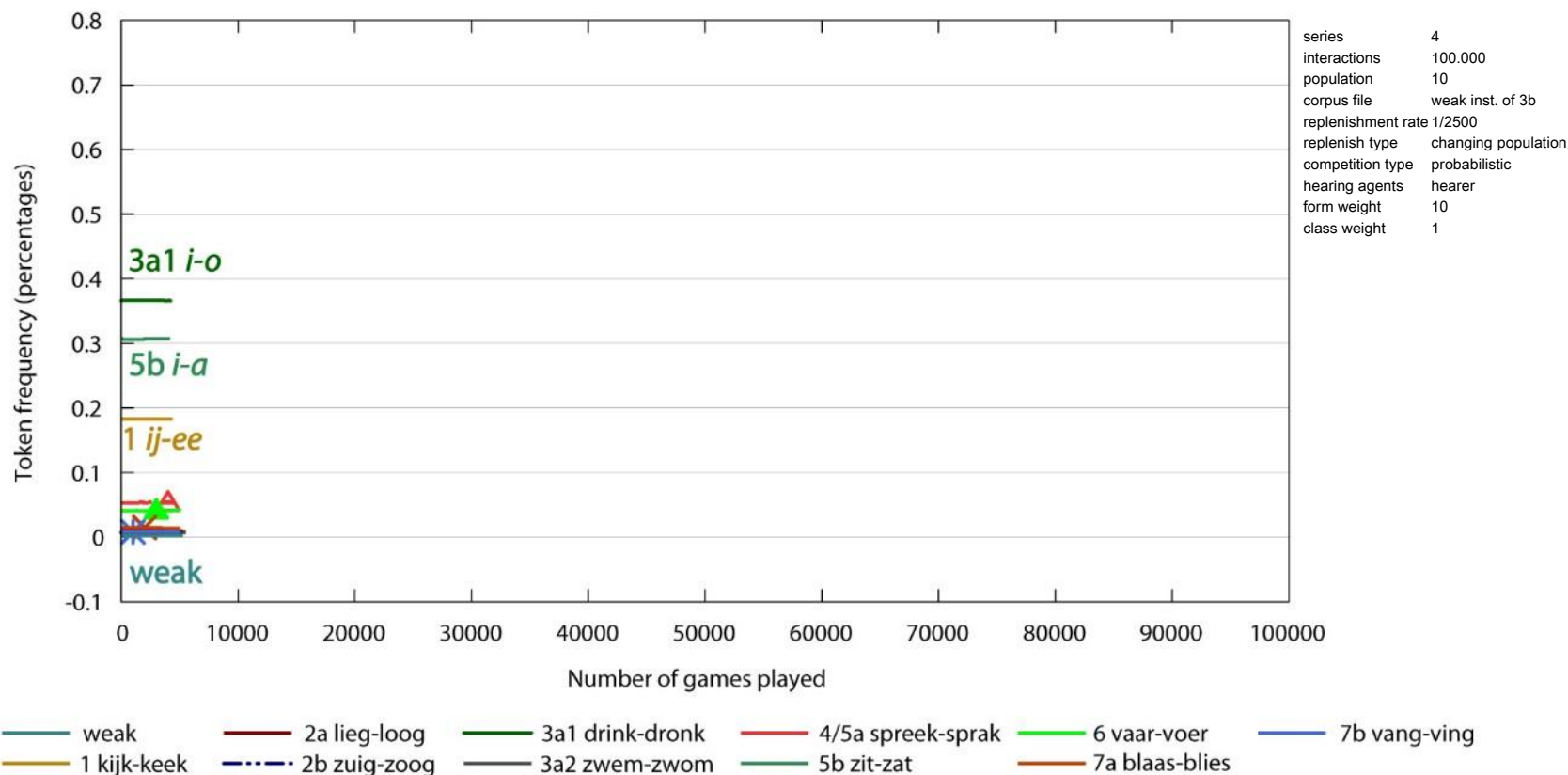
The Results

- Original corpus input, new agent every 2500 interactions



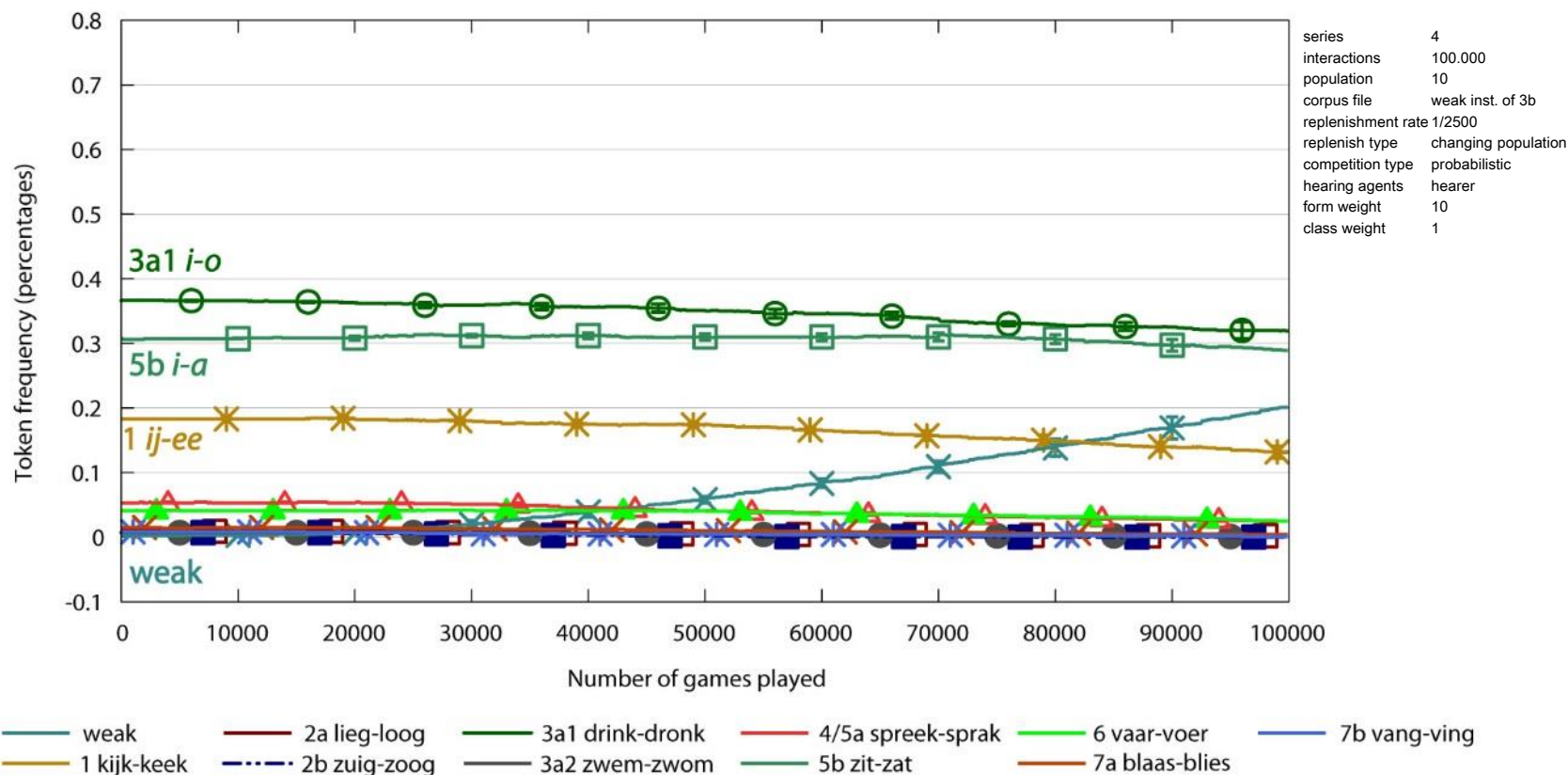
The Results

- Weak starting from vastly inferior position, new agent every 2500 interactions



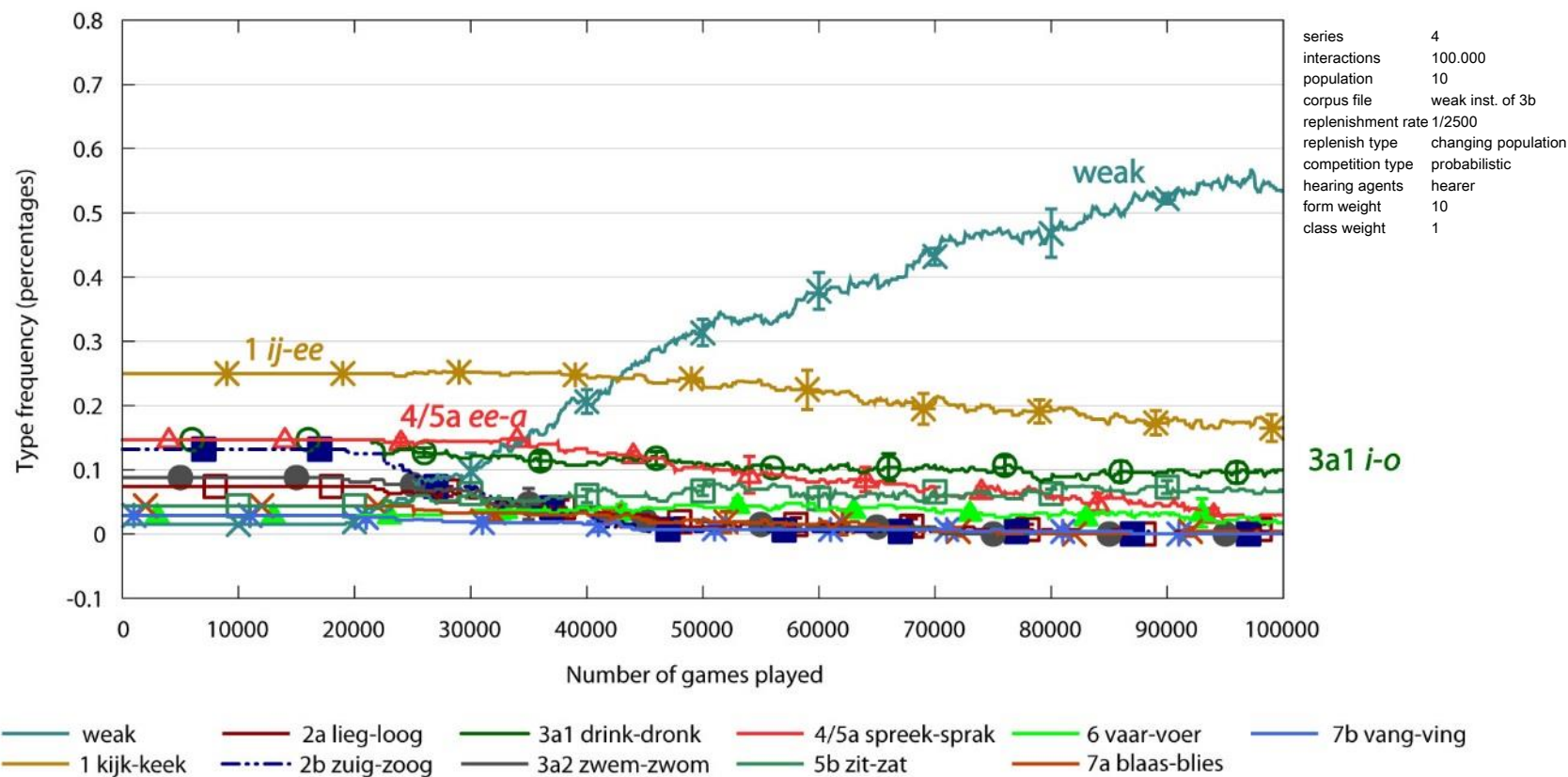
The Results

- Weak starting from vastly inferior position, new agent every 2500 interactions



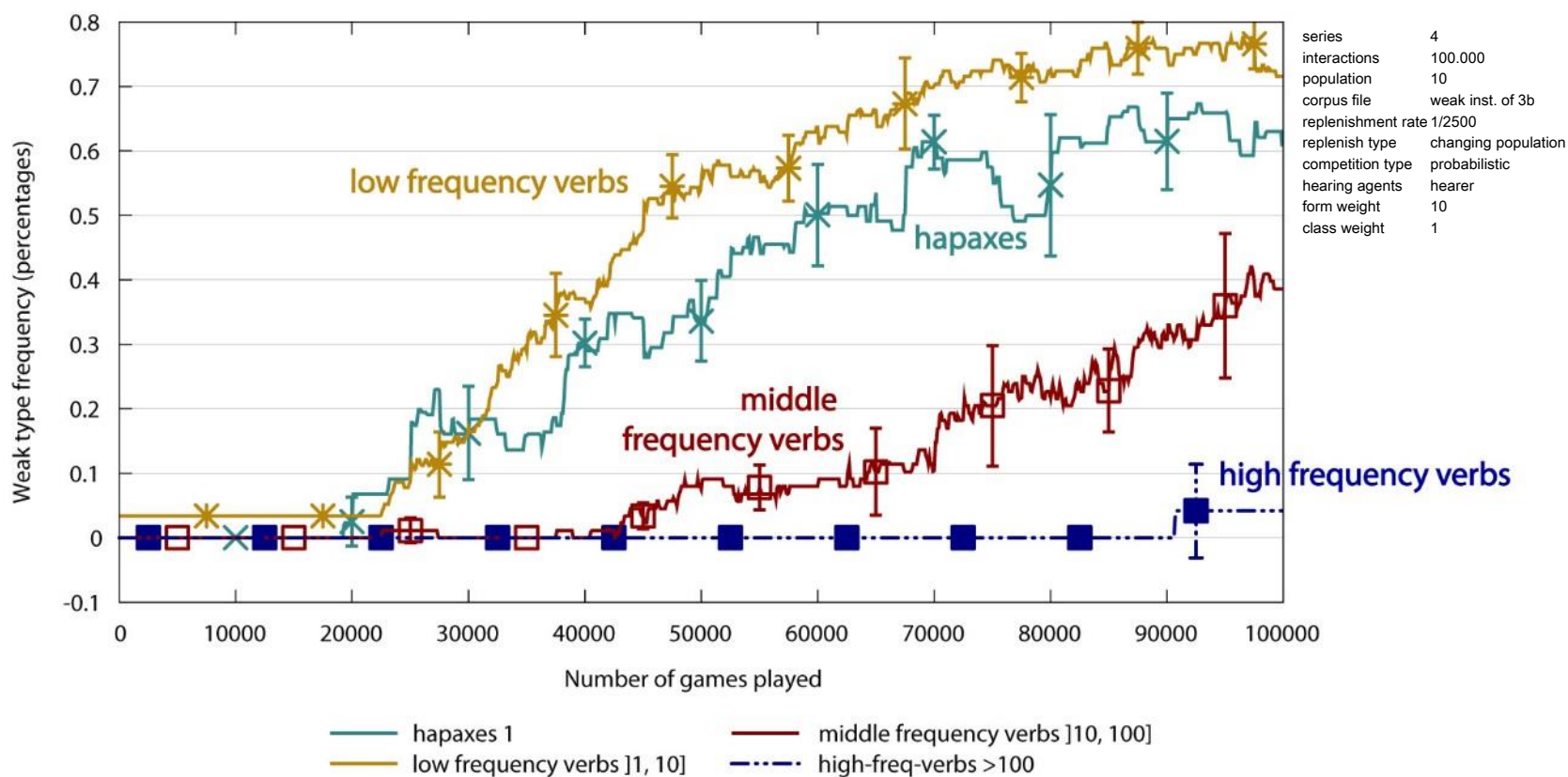
The Results

- Weak starting from vastly inferior position, new agent every 2500 interactions



The Results

- Weak starting from vastly inferior position, new agent every 2500 interactions



The Conclusions

- General applicability as an advantage to learners
- No short term effect
- Given the right sociohistorical conditions, huge long term effect

Thanks

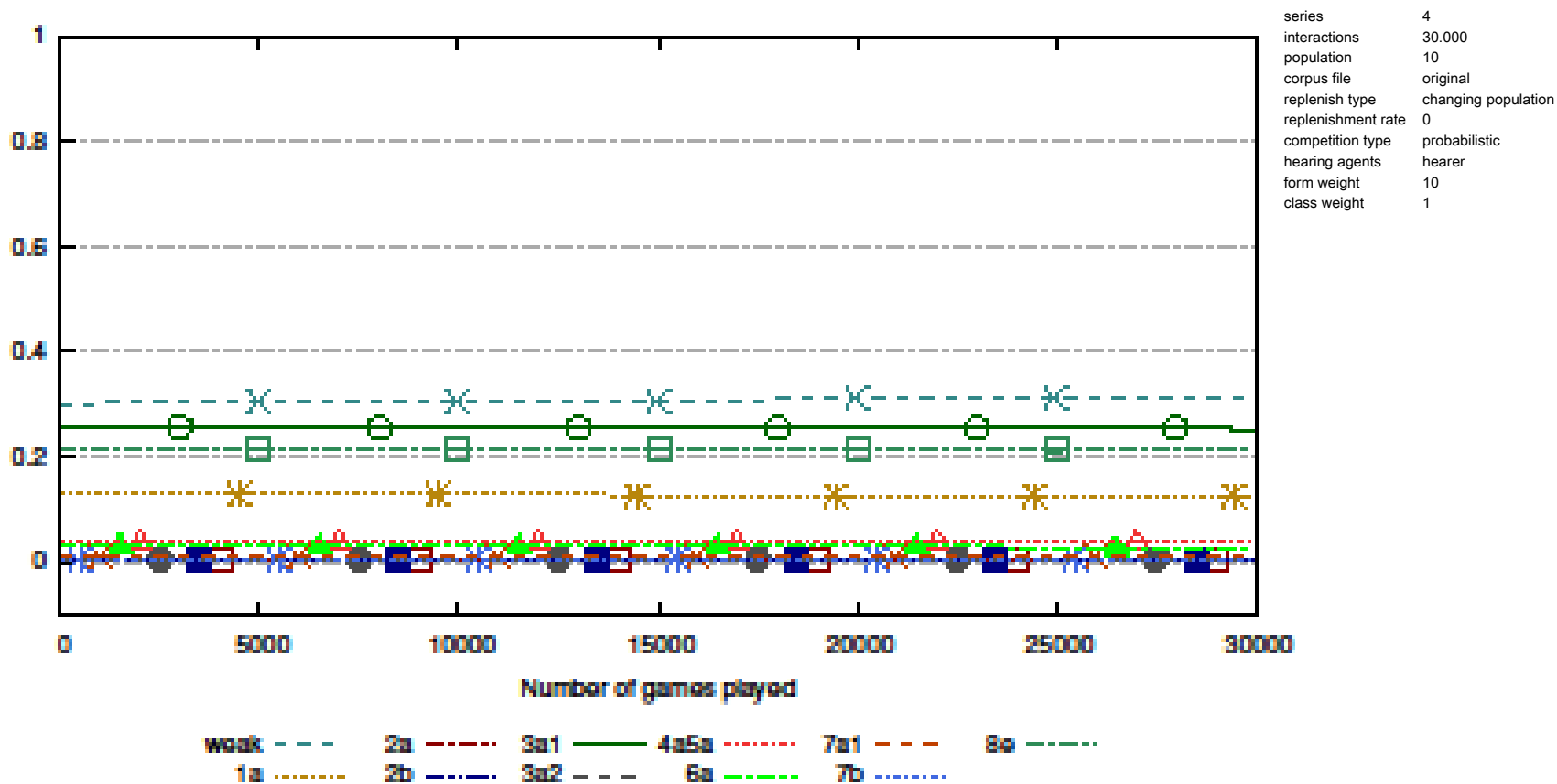
- for further information: dirk.pijpops@kuleuven.be

References

- Carroll, Ryan, Ragnar Svare & Joseph Salmons. 2012. Quantifying the evolutionary dynamics of German verbs. (ed.) Ryan Carroll. *Journal of Historical Linguistics* 2(2). 153–172.
- Grimm, Jacob. 1819. *Deutsche Grammatik*. Göttingen: Dieterichsche Buchhandlung.
- Lieberman, E., Michel, J.-B., Jackson, B., Tang, T., & A. Martin. 2007. Quantifying the evolutionary dynamics of language. *Nature* 449 (7163): 713–716.
- Pijpops, Dirk, Katrien Beuls, Pieter Wellens & Freek Van de Velde. 2014. Entrenchment vs. transparency. Modelling the Dutch strong-weak past tense competition in an agent-based simulation. *Workshop poster presentation at Evolang X, 10th International Conference on the Evolution of Language, April 14-17, 2014, Department of English of Vienna University, Vienna, Austria*.
- Pugliesi, Martina, Vittorio Loreto, Christine Cuskley, Claudio Castellano & Franscesca Colaiori. 2014. The role of coordination in regularization. In Erica A. Cartmill, Seán Roberts, Heidi Lyn & Hannah Cornish (eds.), *The Evolution of Language: Proceedings of the 10th International Conference*, 260–261. London: World Scientific.

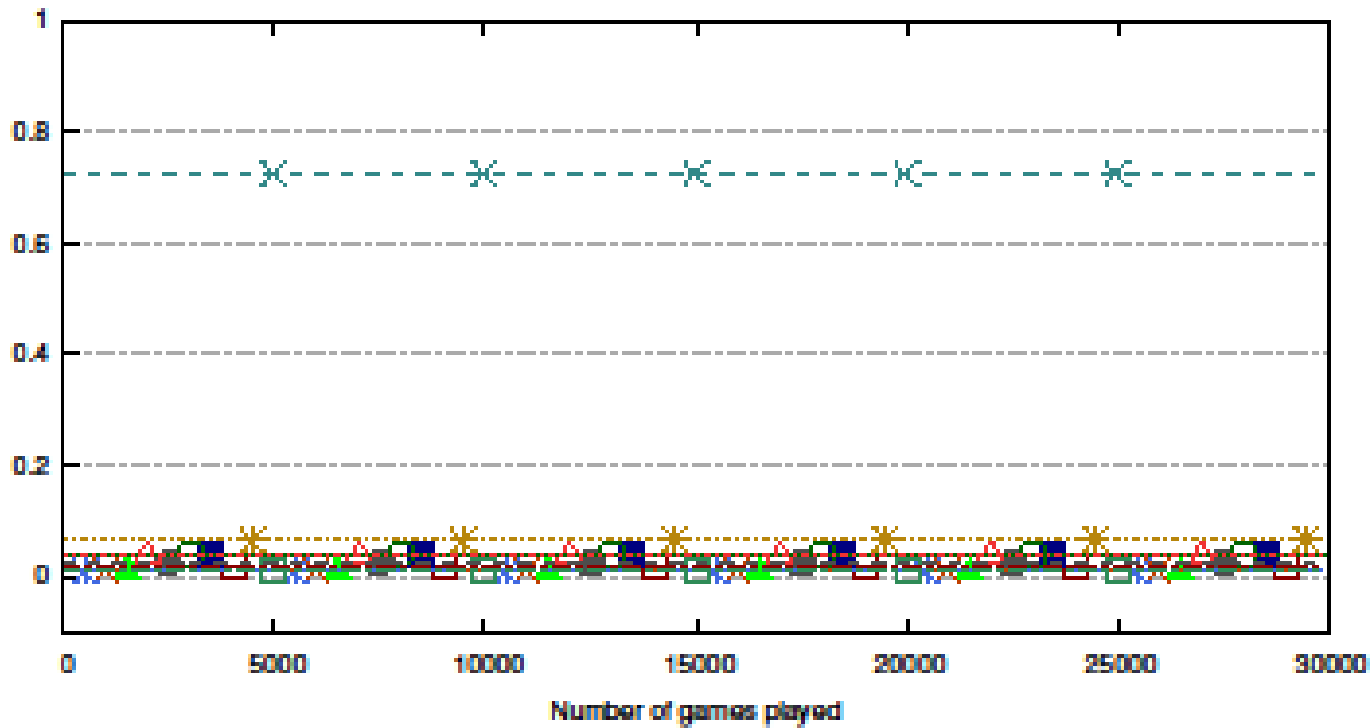
Results

- 1: token frequencies



Results

- 1: type frequencies

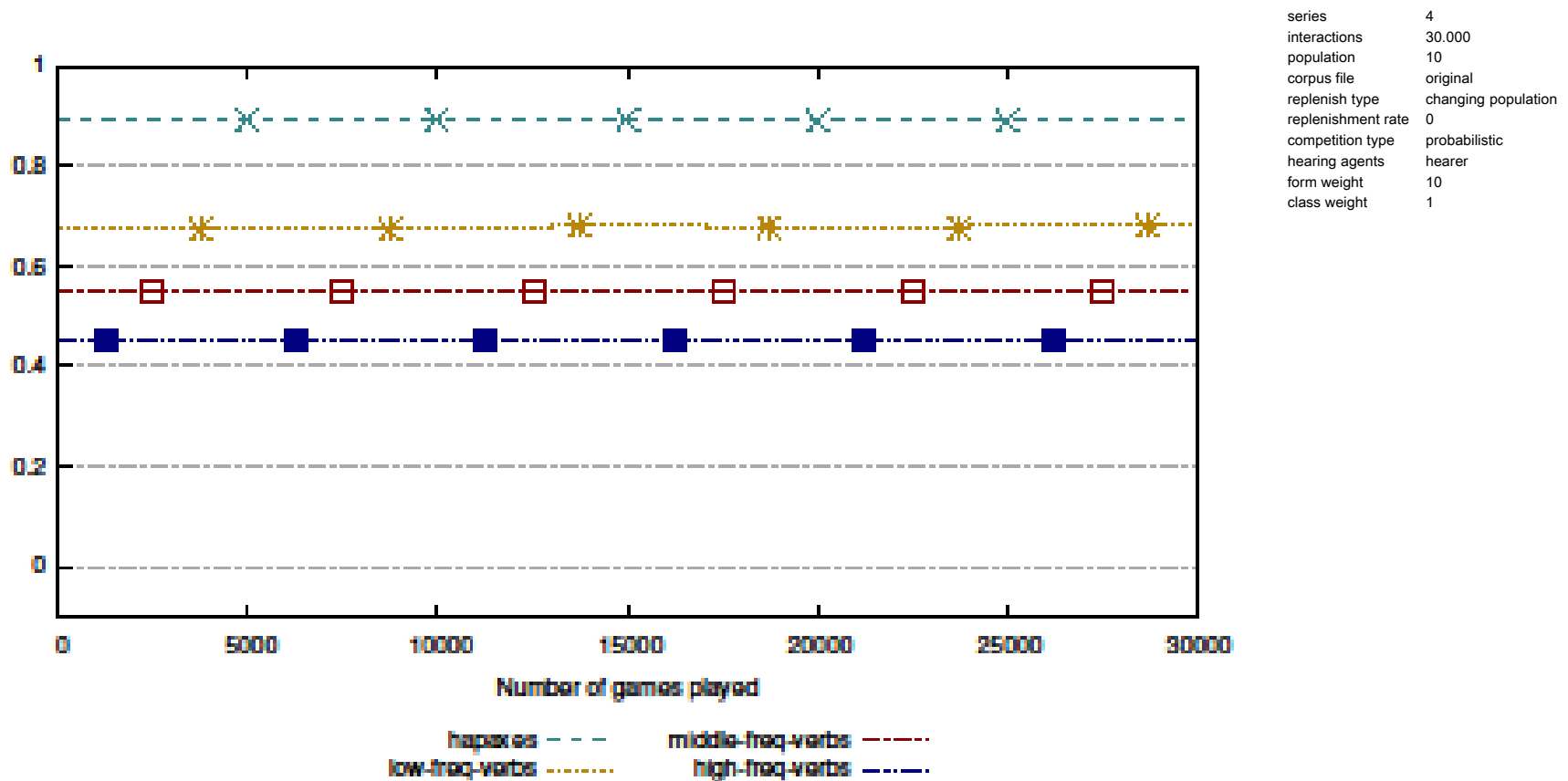


series	4
interactions	30.000
population	10
corpus file	original
replenish type	changing population
replenishment rate	0
competition type	probabilistic
hearing agents	hearer
form weight	10
class weight	1

weak - - - 2a - - - 3a1 - - - 4a5a - - - 7a1 - - - 8a - - -
 1a - - - 2b - - - 3a2 - - - 6a - - - 7b - - -

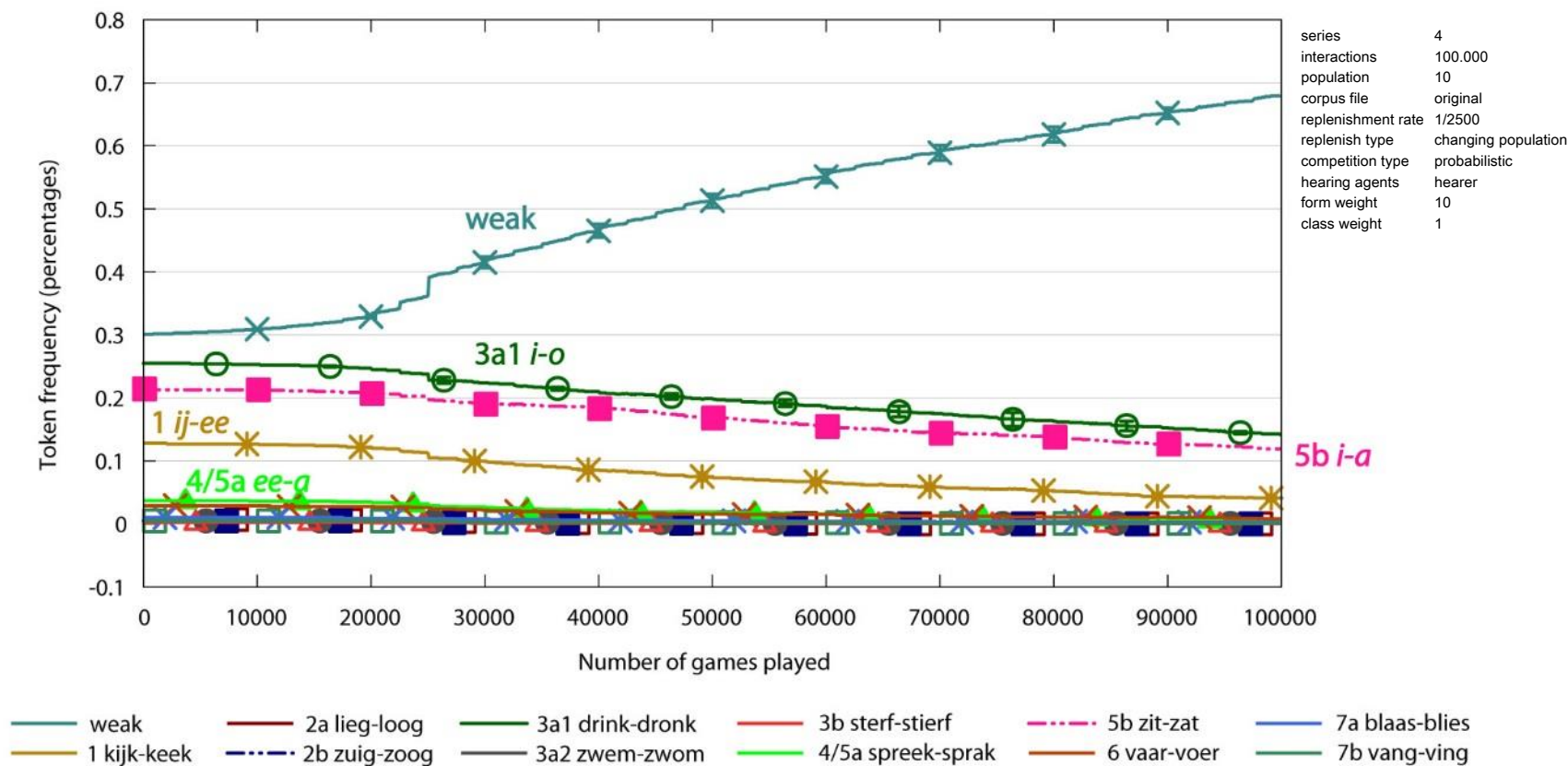
Results

- 1: frequency bins



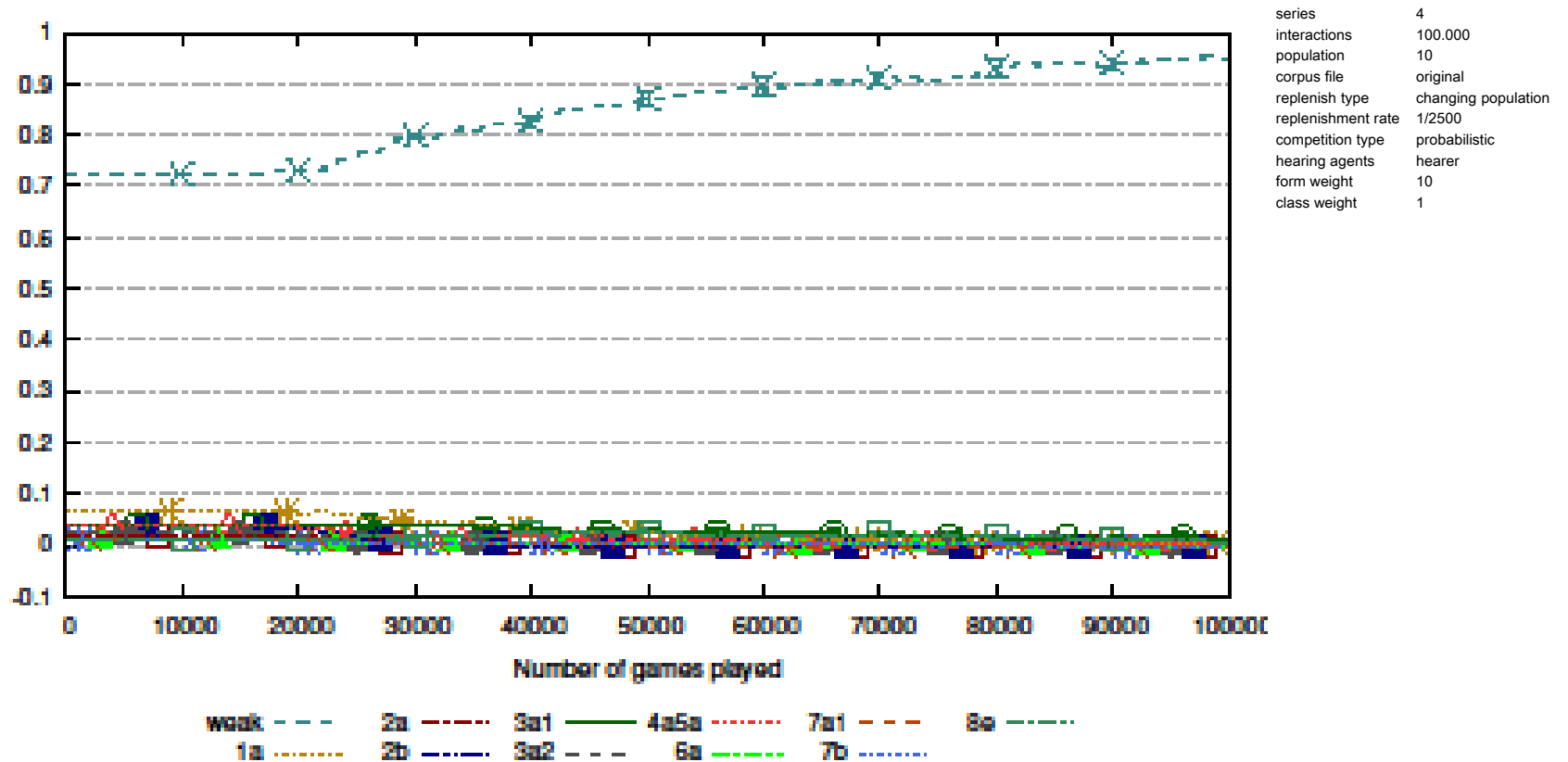
The Results

- 2: token frequencies



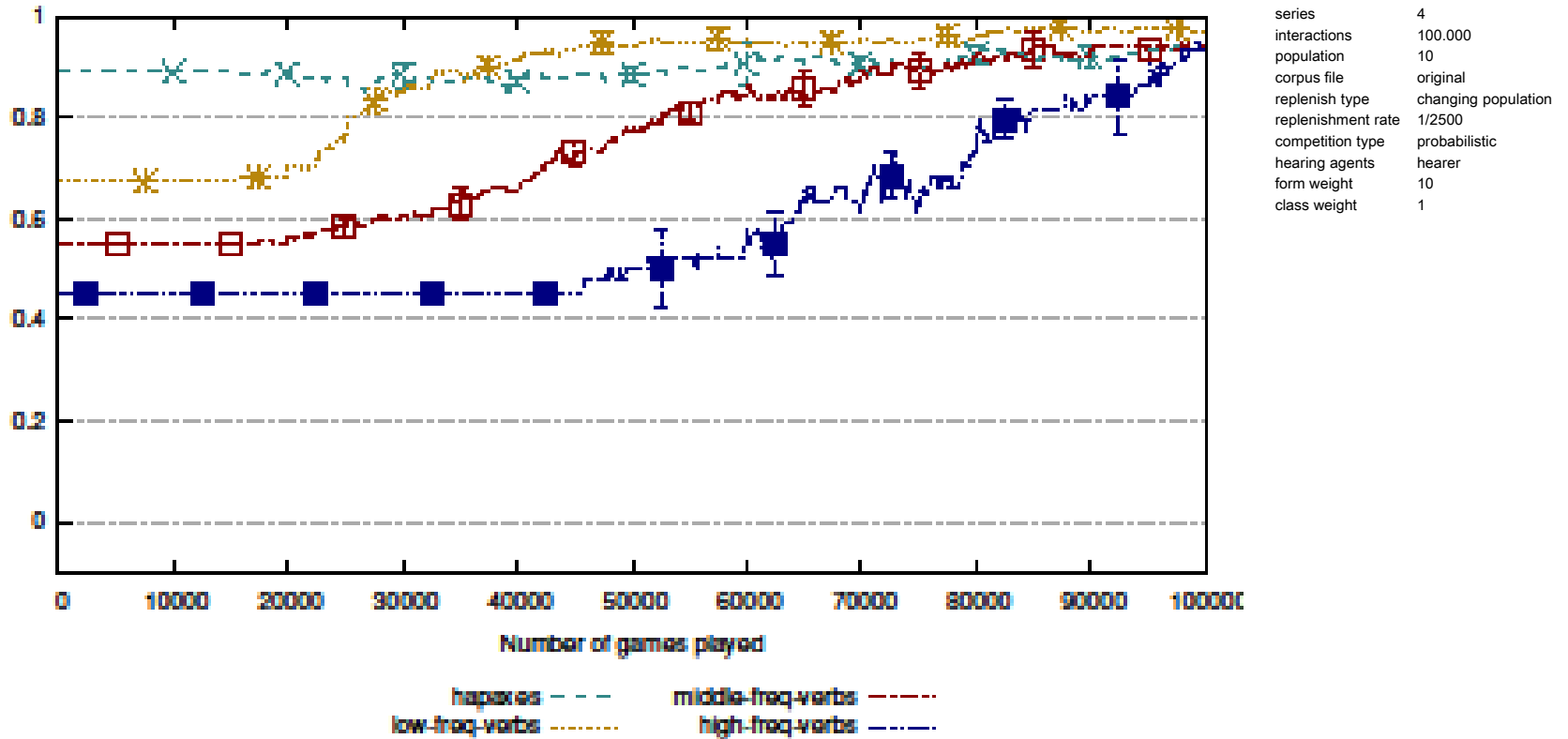
Results

- 2: type frequencies



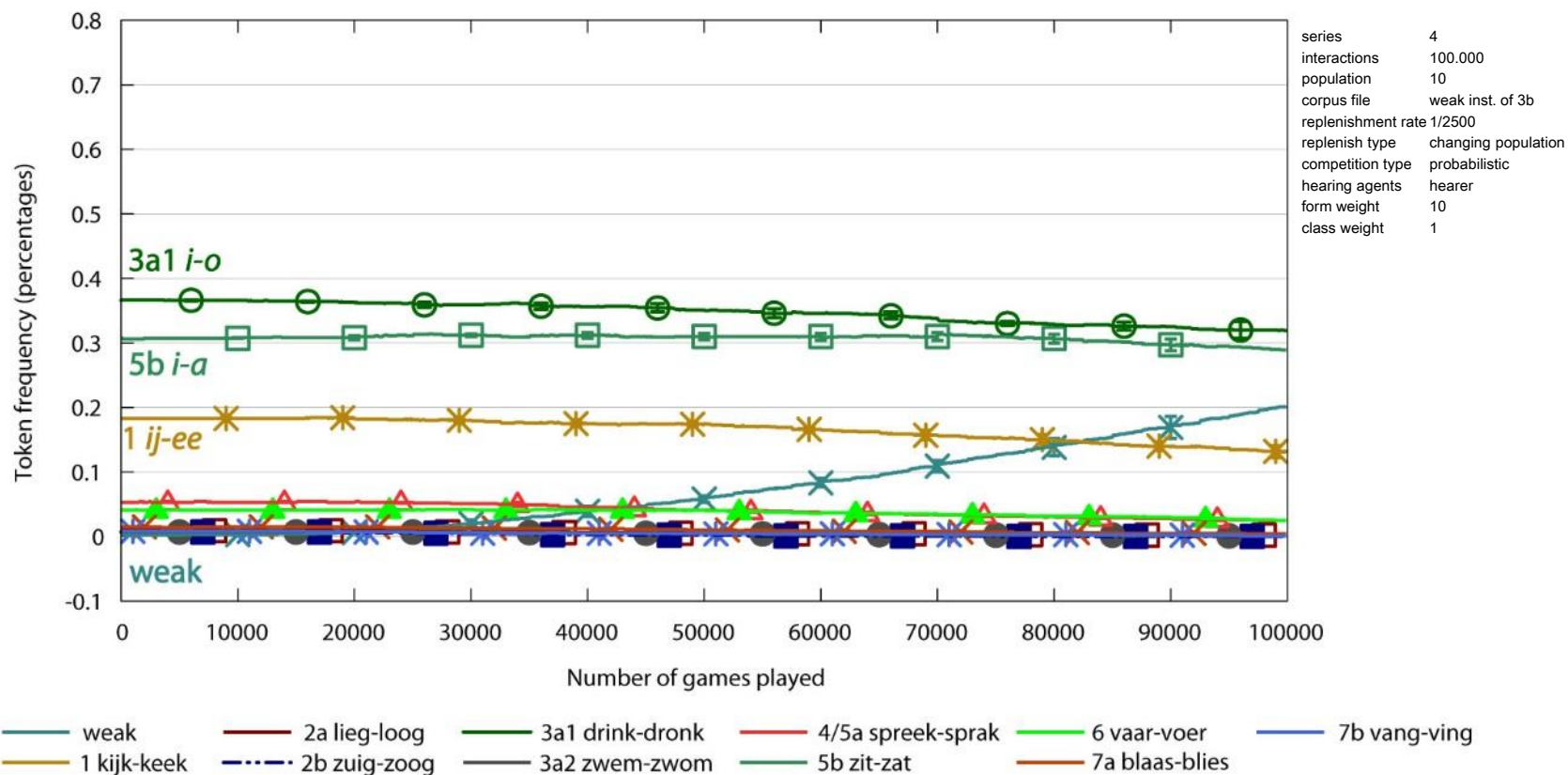
Results

- 2: frequency bins



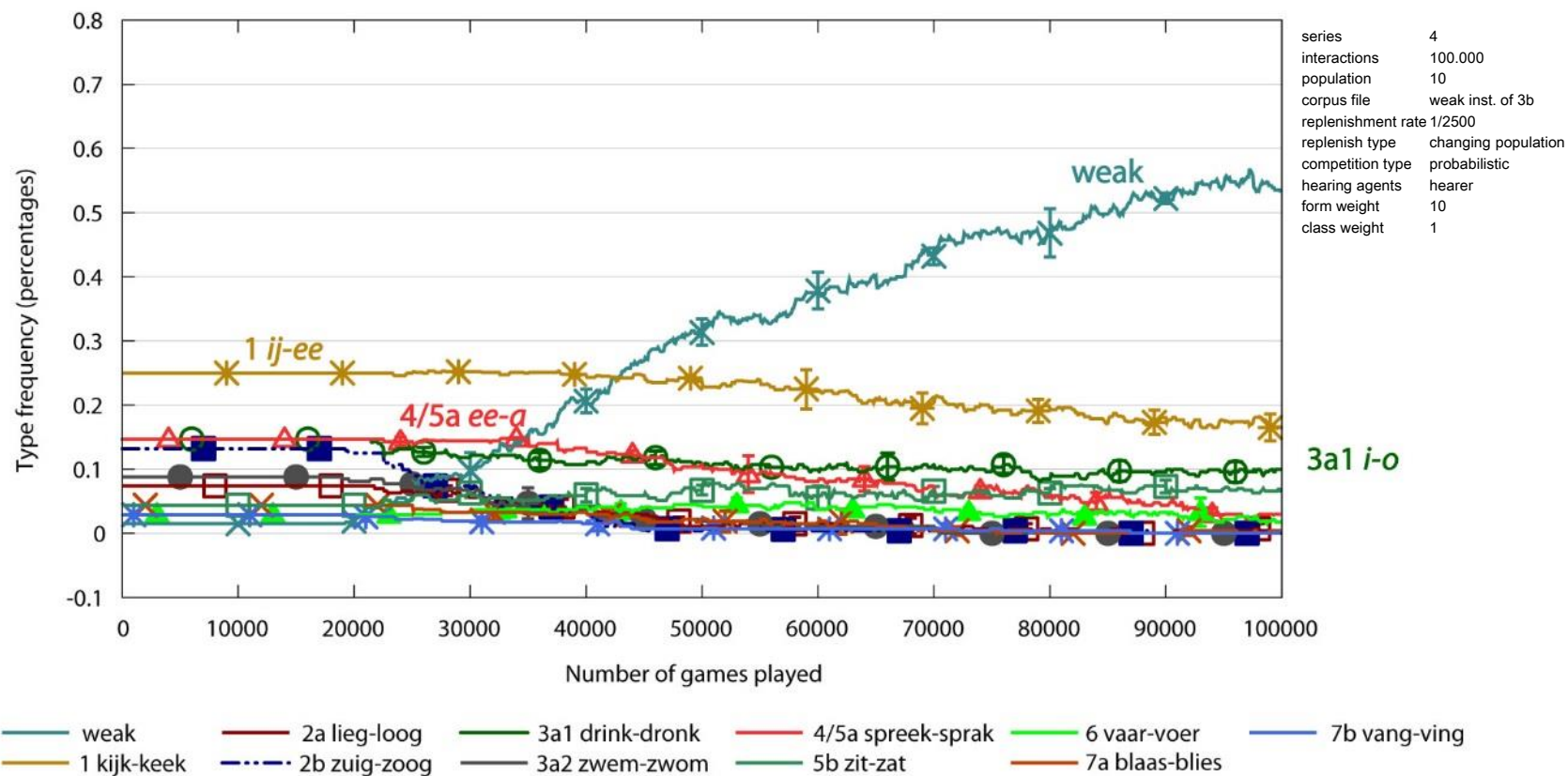
The Results

- 3: token frequencies



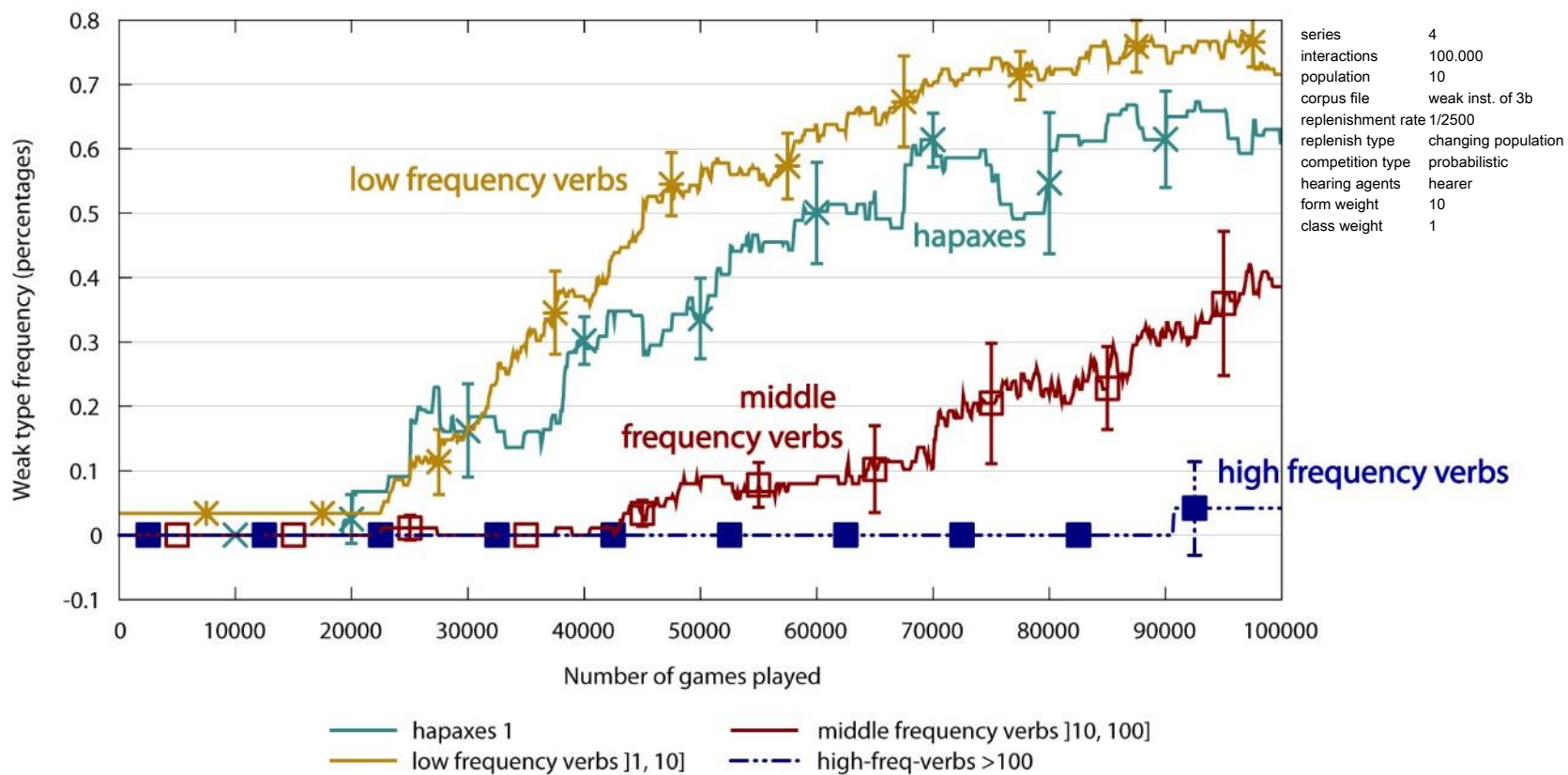
The Results

- 3: type frequencies



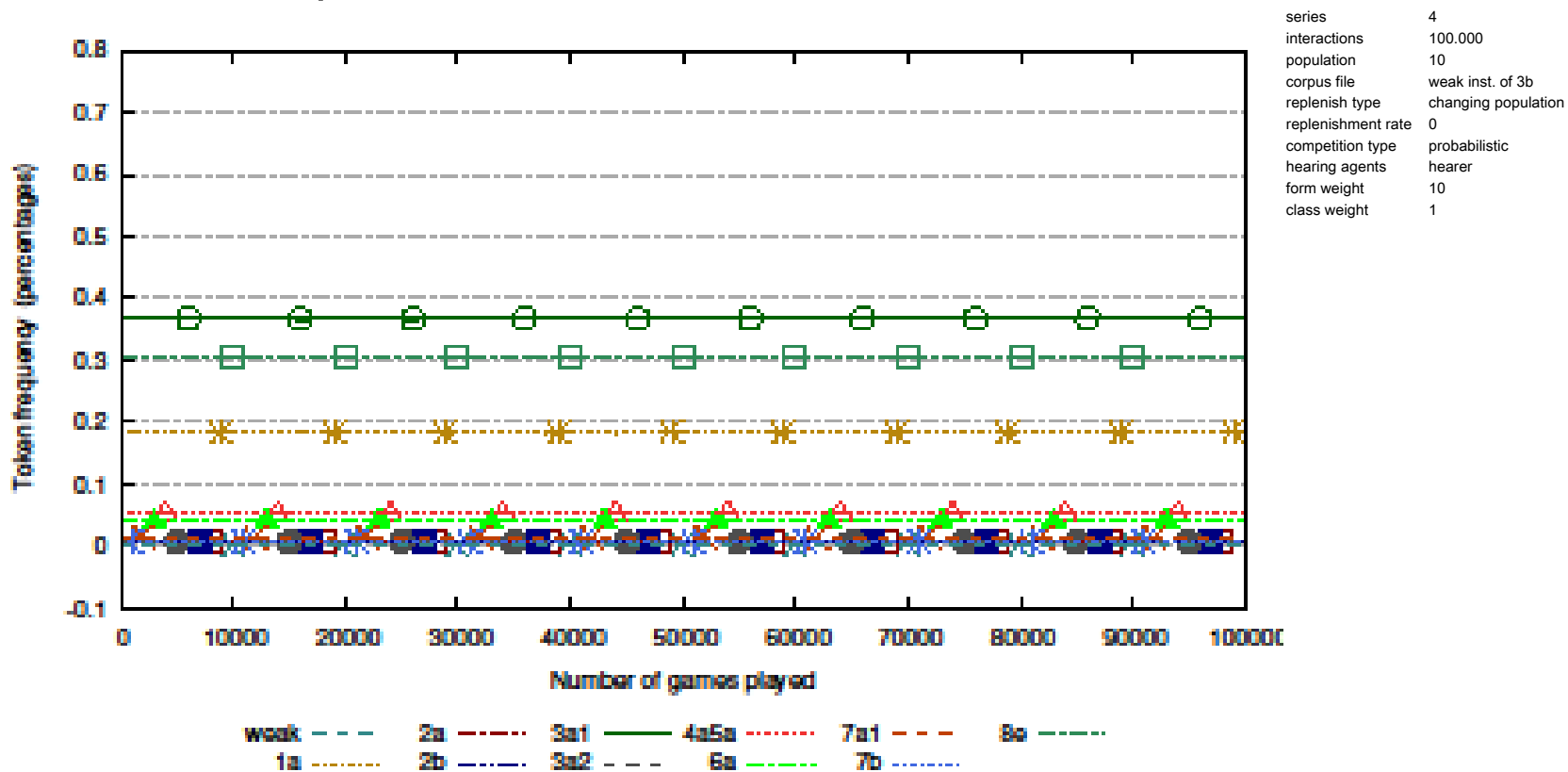
The Results

- 3: frequency bins



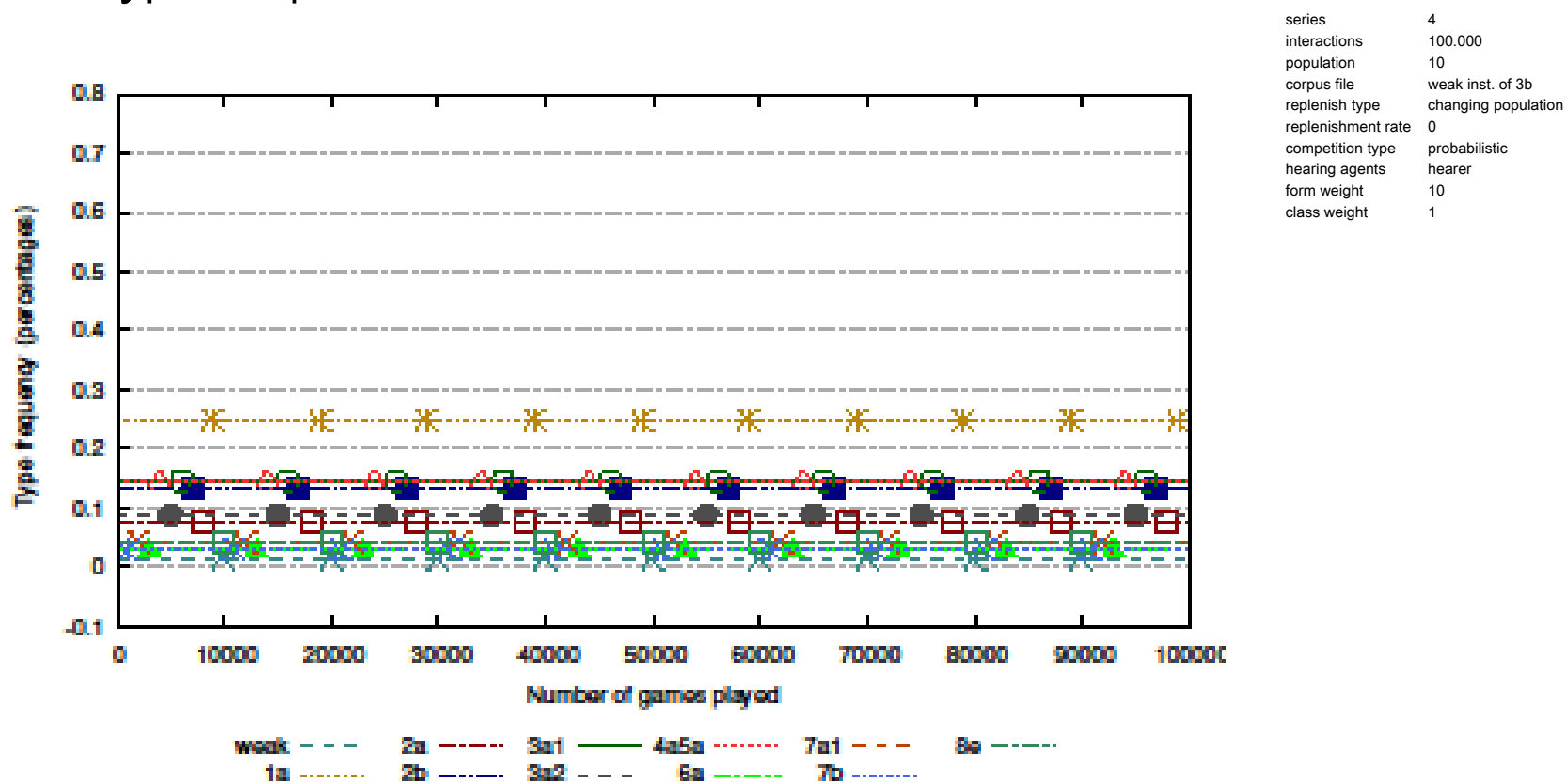
Results

- 4: token frequencies



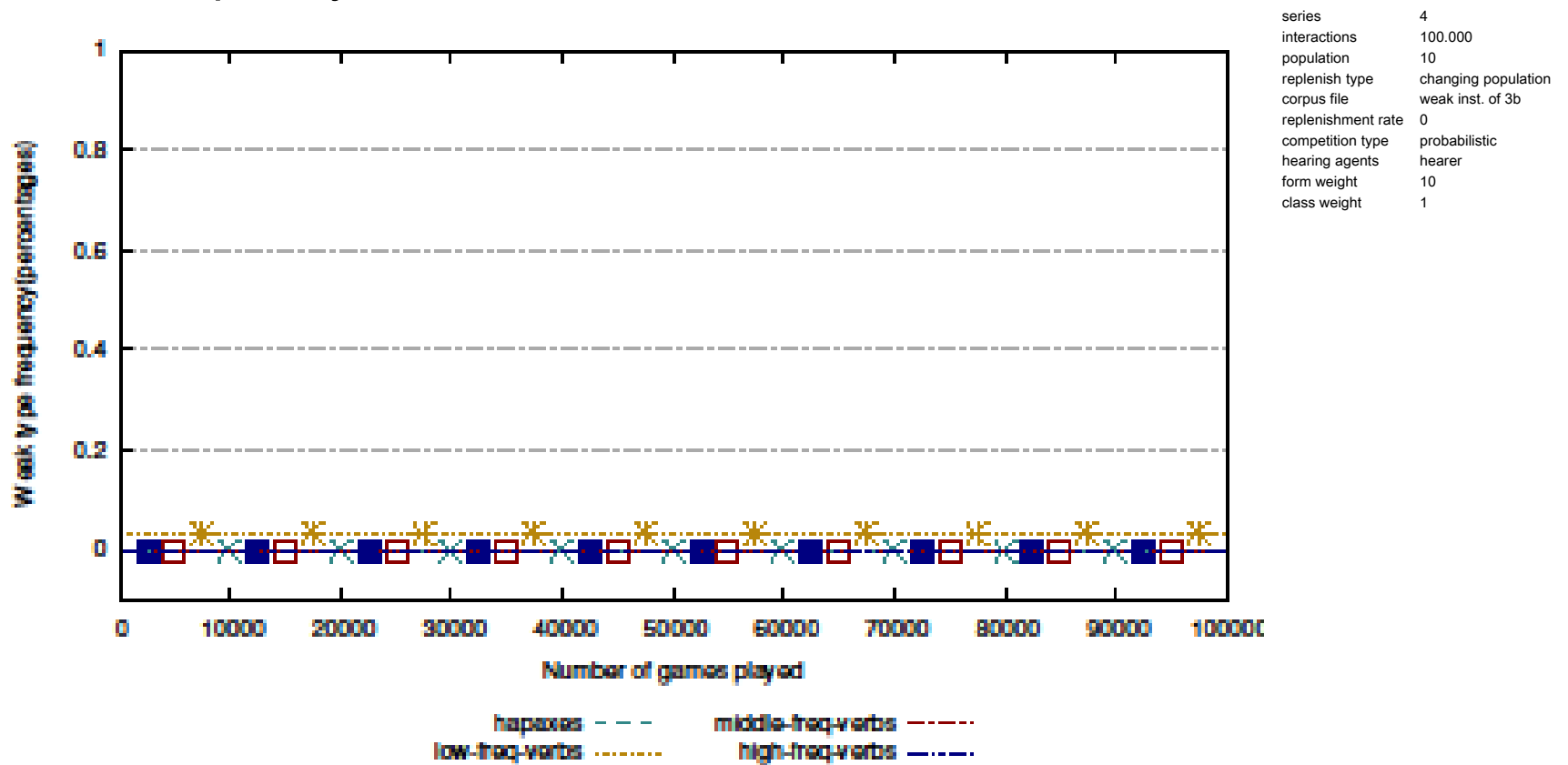
Results

- 4: type frequencies



Results

- 4: frequency bins

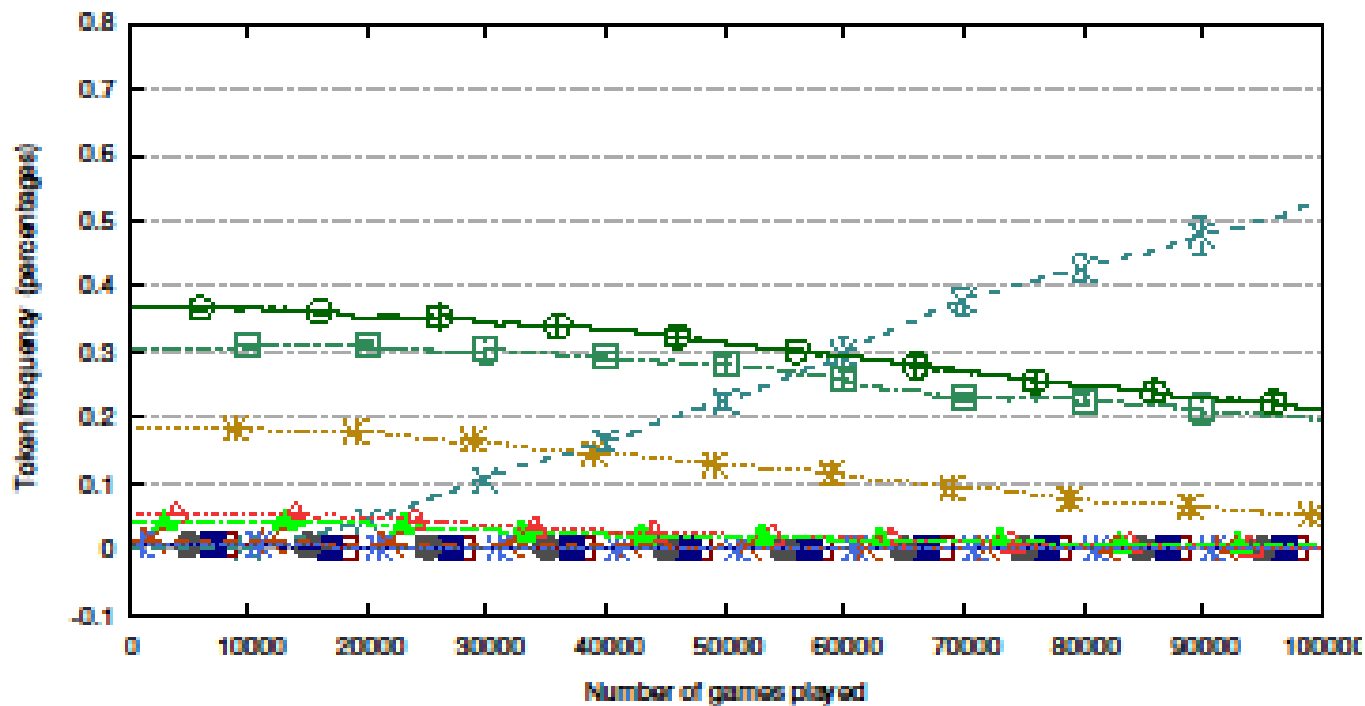


Results

- 5: token frequencies

```

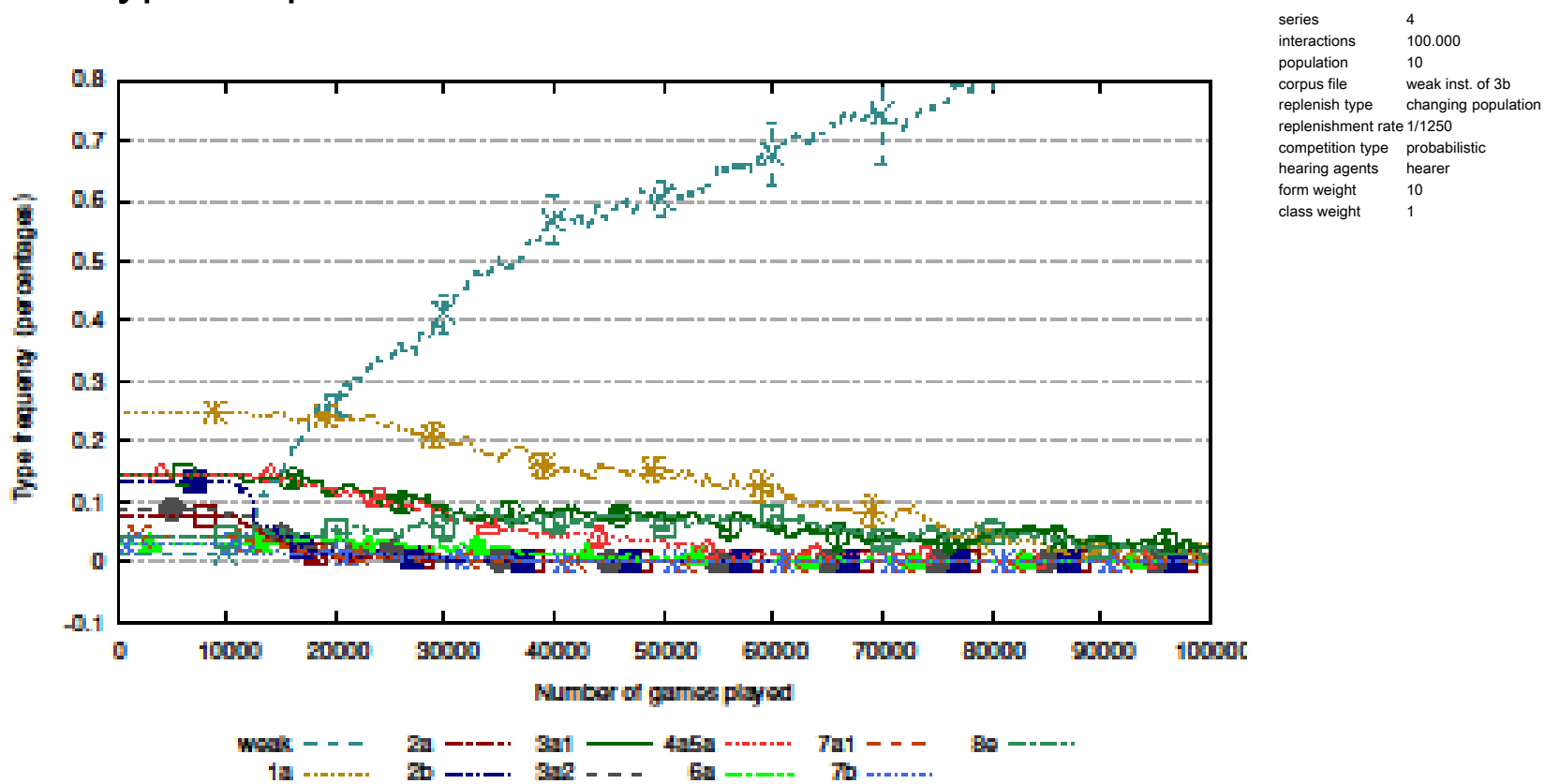
series      4
interactions 100.000
population  10
corpus file  weak inst. of 3b
replenish type  changing population
replenishment rate 1/1250
competition type  probabilistic
hearing agents  hearer
form weight  10
class weight  1
    
```



weak --- 2a --- 3a1 --- 4a5a --- 7a1 --- 8a ---
 1a - - - 2b - - - 3a2 - - - 6a - - - 7b - - -

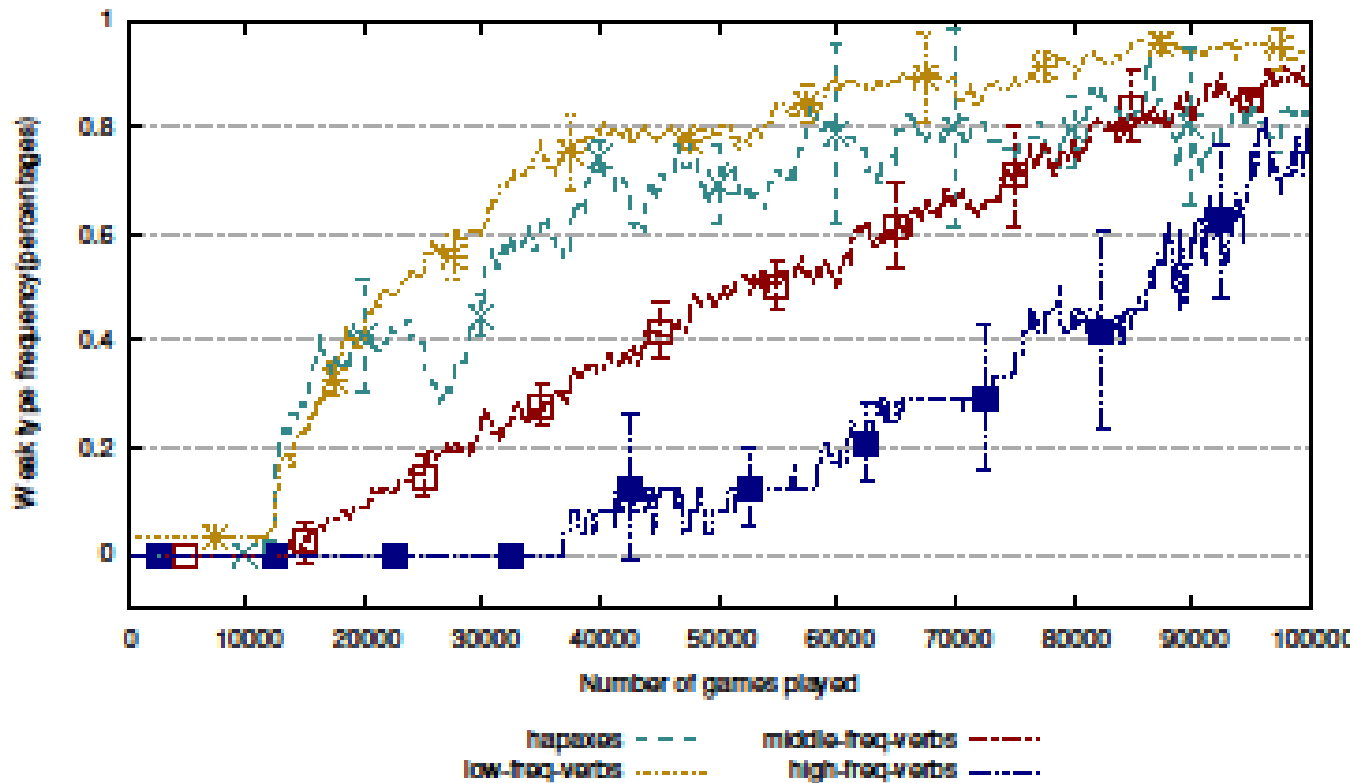
Results

- 5: type frequencies



Results

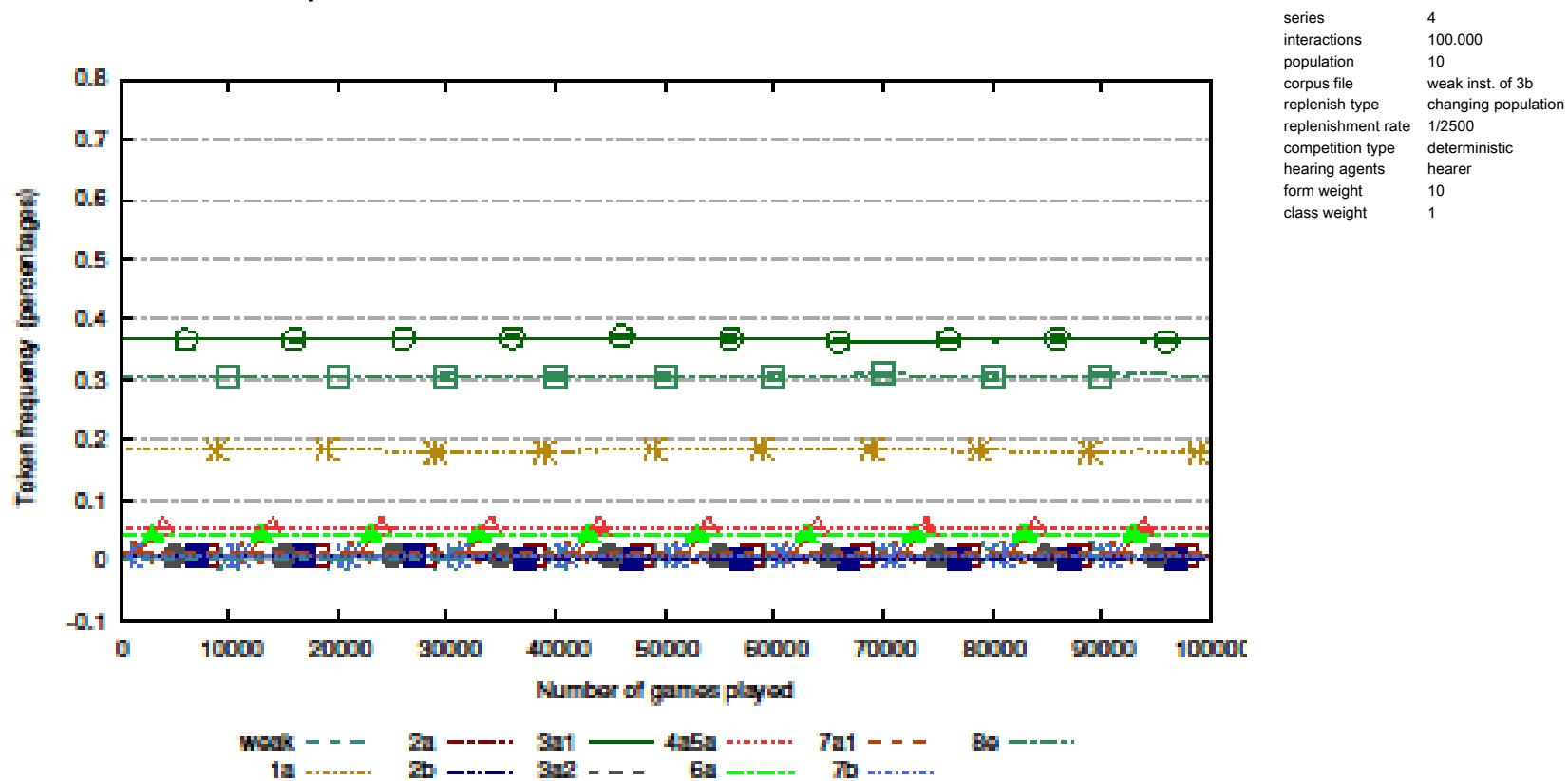
- 5: frequency bins



series	4
interactions	100.000
population	10
corpus file	weak inst. of 3b
replenish type	changing population
replenishment rate	1/1250
competition type	probabilistic
hearing agents	hearer
form weight	10
class weight	1

Results

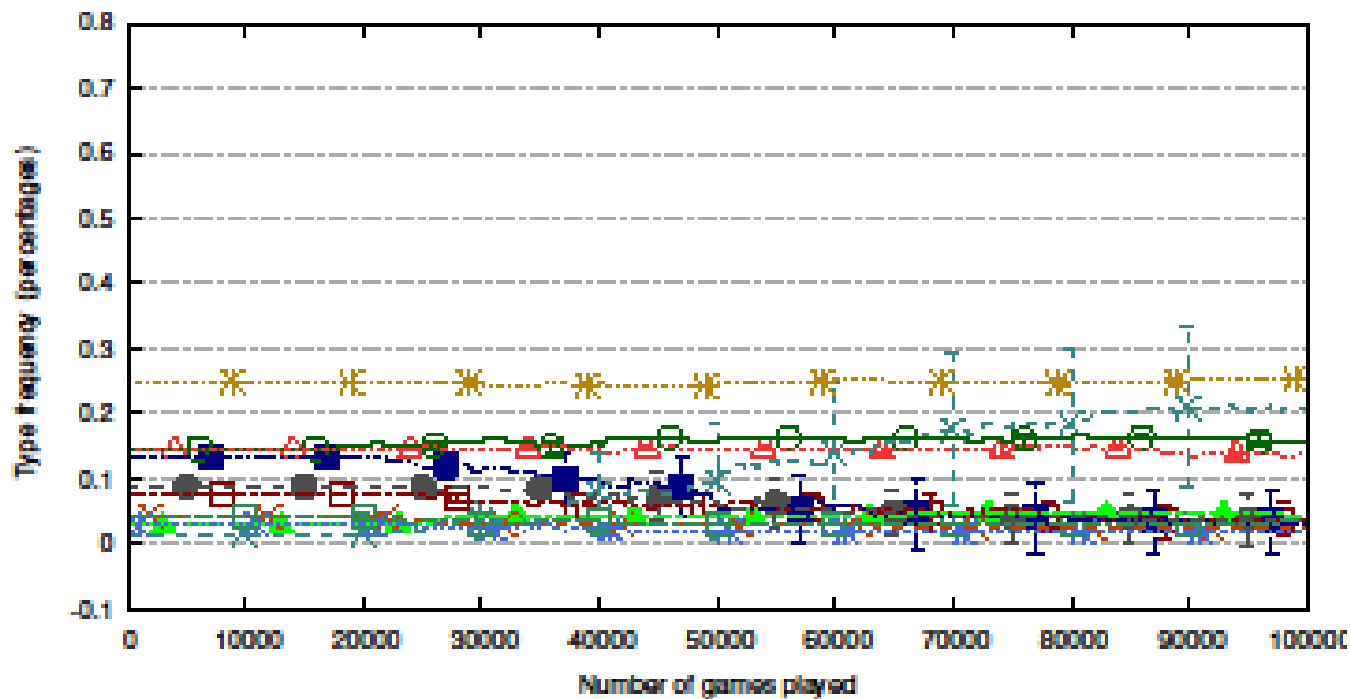
- 6: token frequencies



Results

- 6: type frequencies

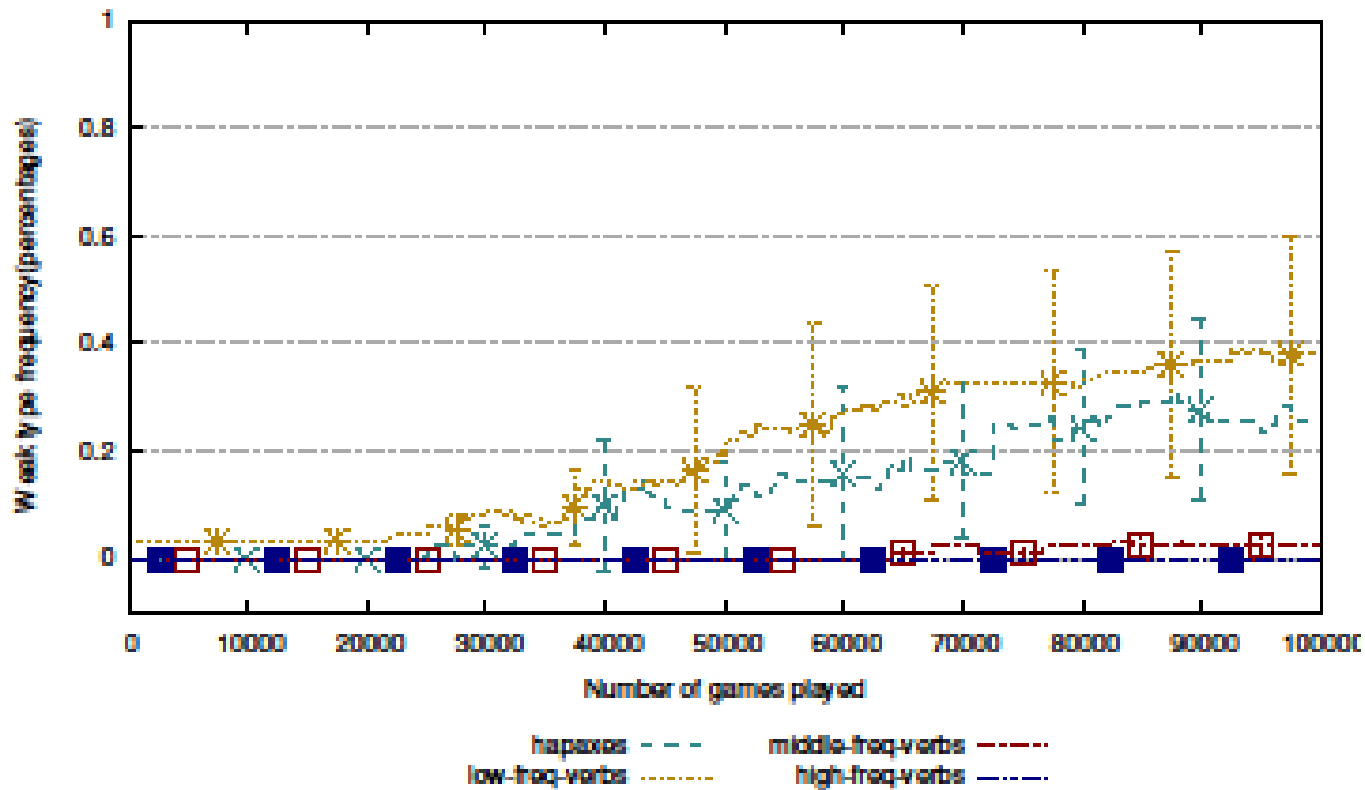
series	4
interactions	100.000
population	10
corpus file	weak inst. of 3b
replenish type	changing population
replenishment rate	1/2500
competition type	deterministic
hearing agents	hearer
form weight	10
class weight	1



weak	2a	3a1	4a5a	7a1	8a
1a	2b	3a2	6a	7b	

Results

- 6: frequency bins



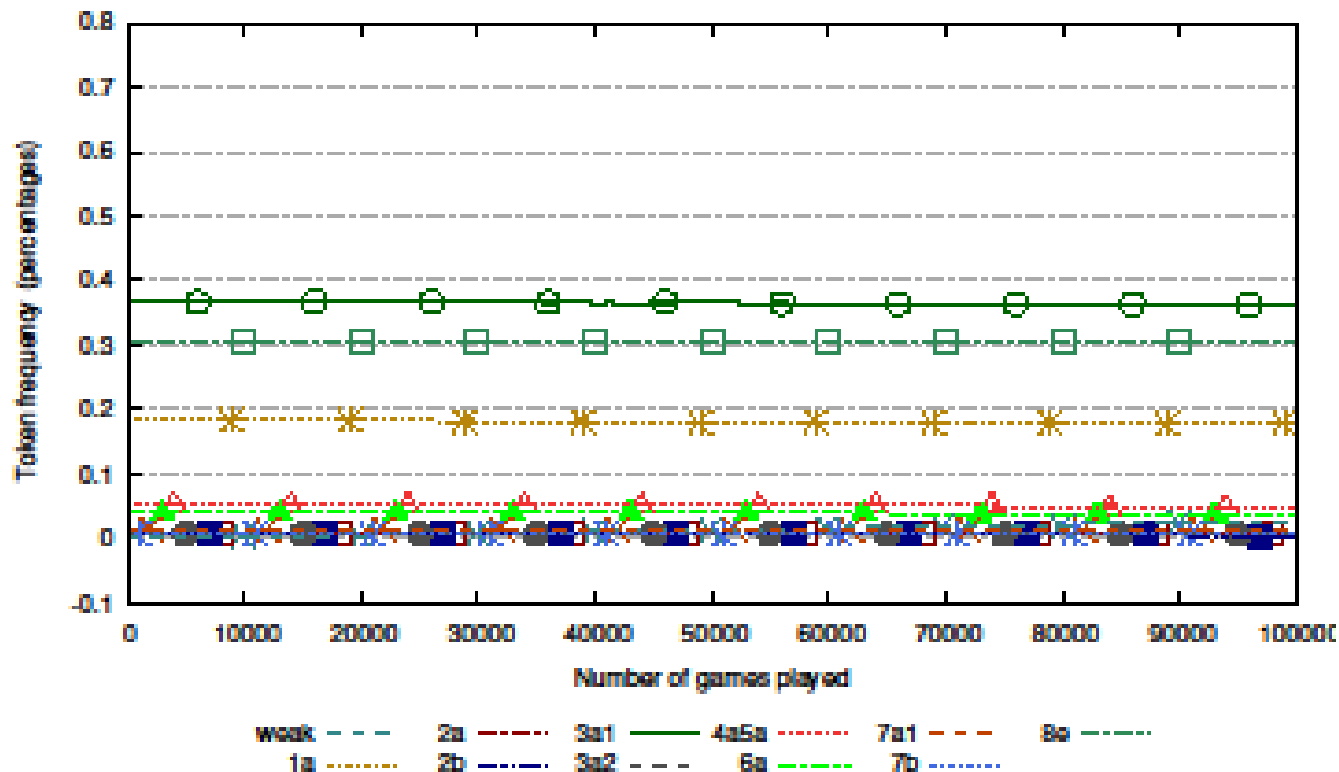
series	4
interactions	100.000
population	10
corpus file	weak inst. of 3b
replenish type	changing population
replenishment rate	1/1250
competition type	probabilistic
hearing agents	hearer
form weight	10
class weight	1

Results

- 7: token frequencies

```

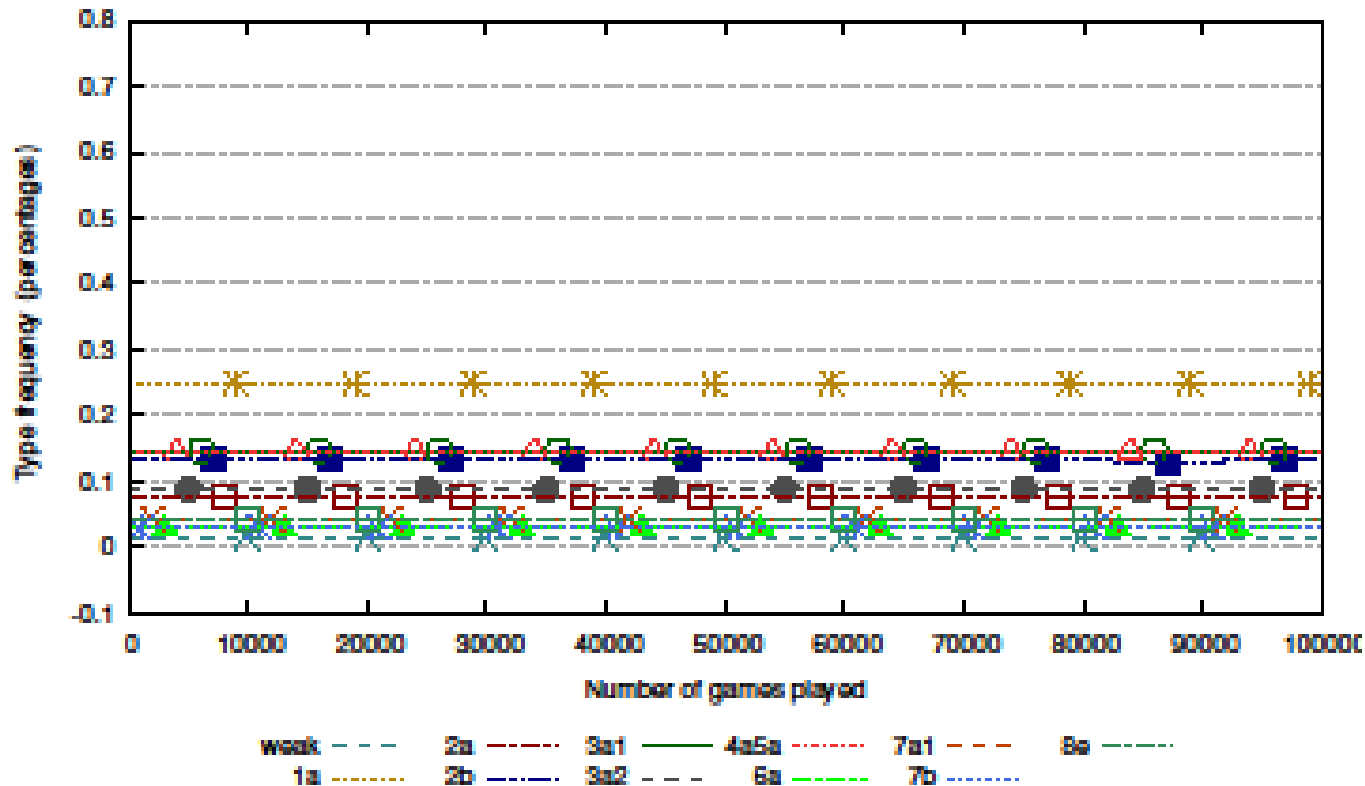
series      4
interactions 100.000
population  10
corpus file  weak inst. of 3b
replenish type  changing population
replenishment rate 1/2500
competition type  probabilistic
hearing agents  hearer
form weight  1.2
class weight  1
    
```



Results

- 7: type frequencies

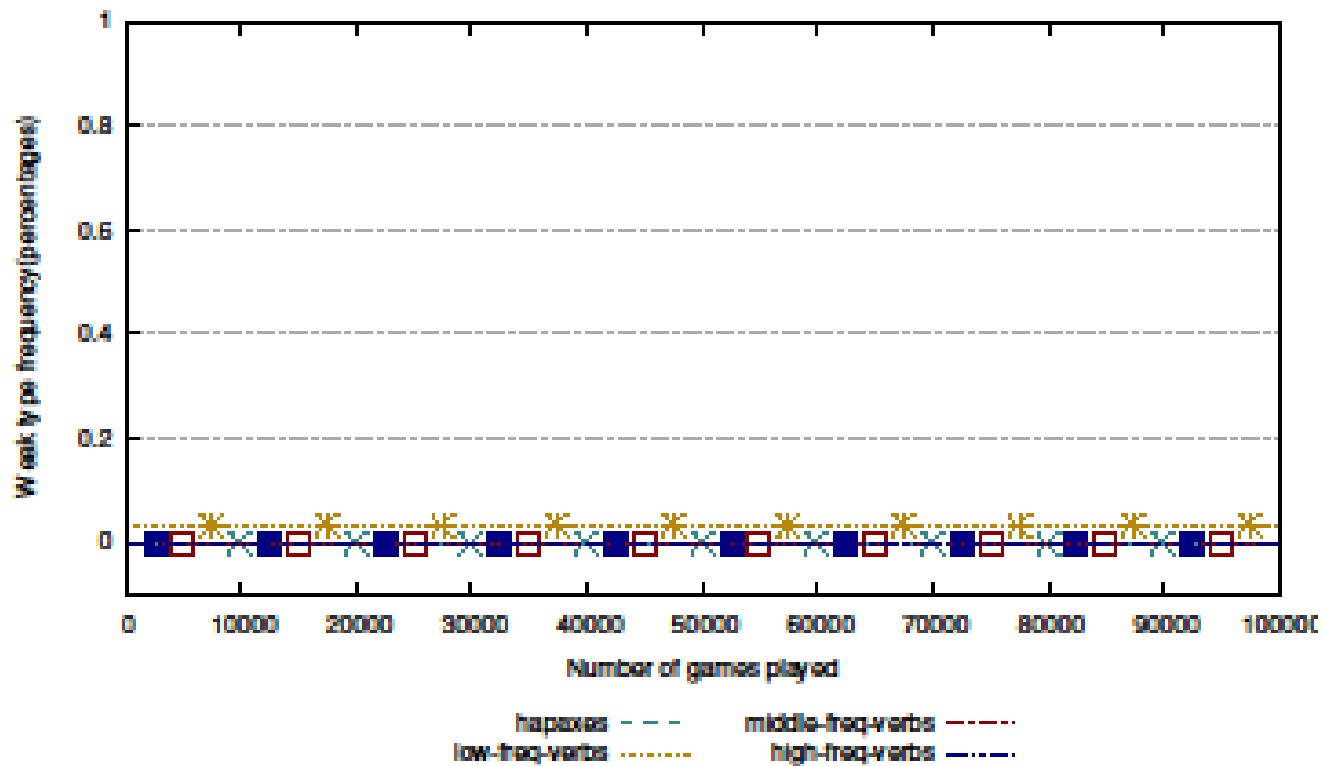
series 4
 interactions 100.000
 population 10
 corpus file weak inst. of 3b
 replenish type changing population
 replenishment rate 1/2500
 competition type probabilistic
 hearing agents hearer
 form weight 1.2
 class weight 1



Results

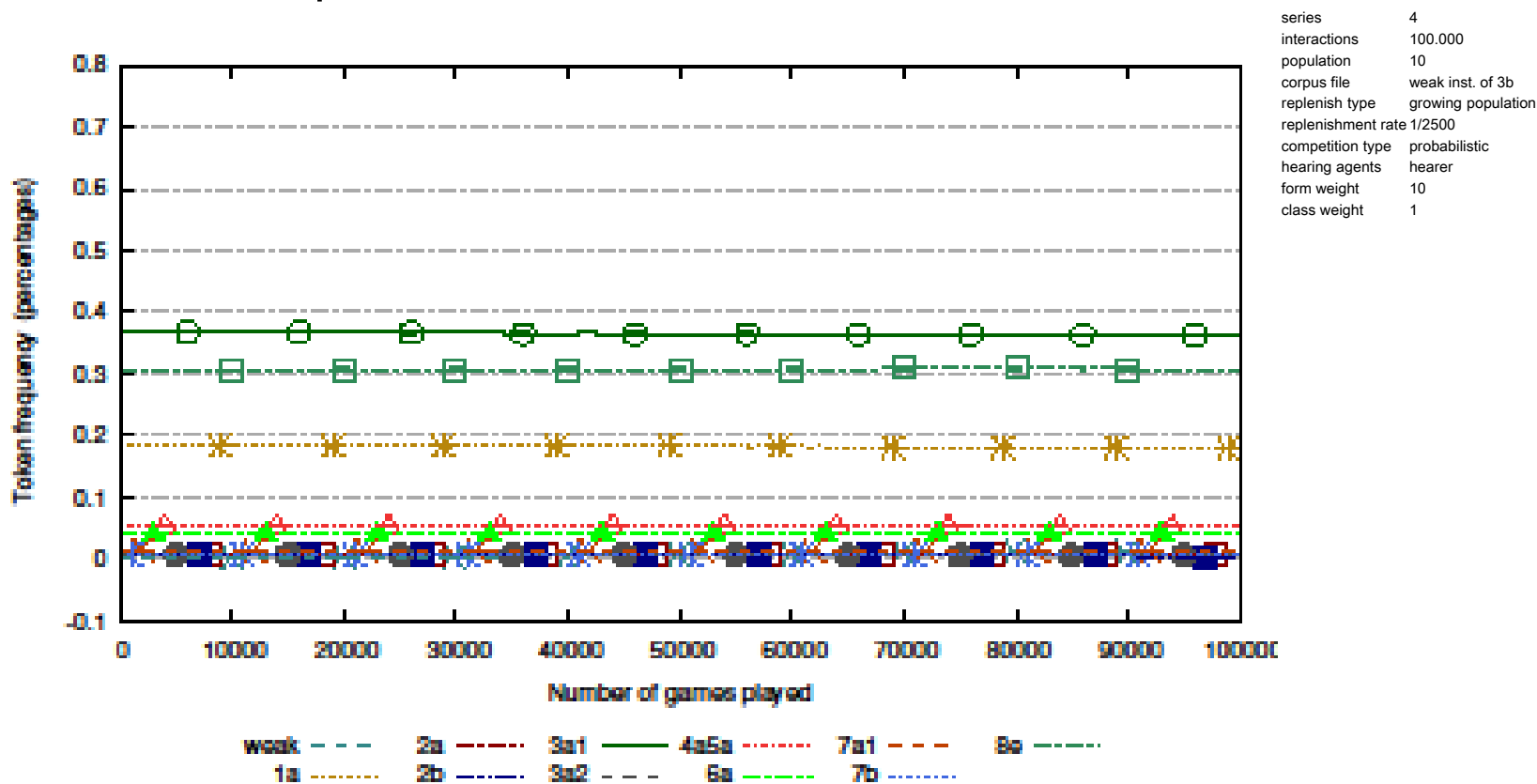
- 7: frequency bins

series	4
interactions	100.000
population	10
corpus file	weak inst. of 3b
replenish type	changing population
replenishment rate	1/2500
competition type	probabilistic
hearing agents	hearer
form weight	1.2
class weight	1



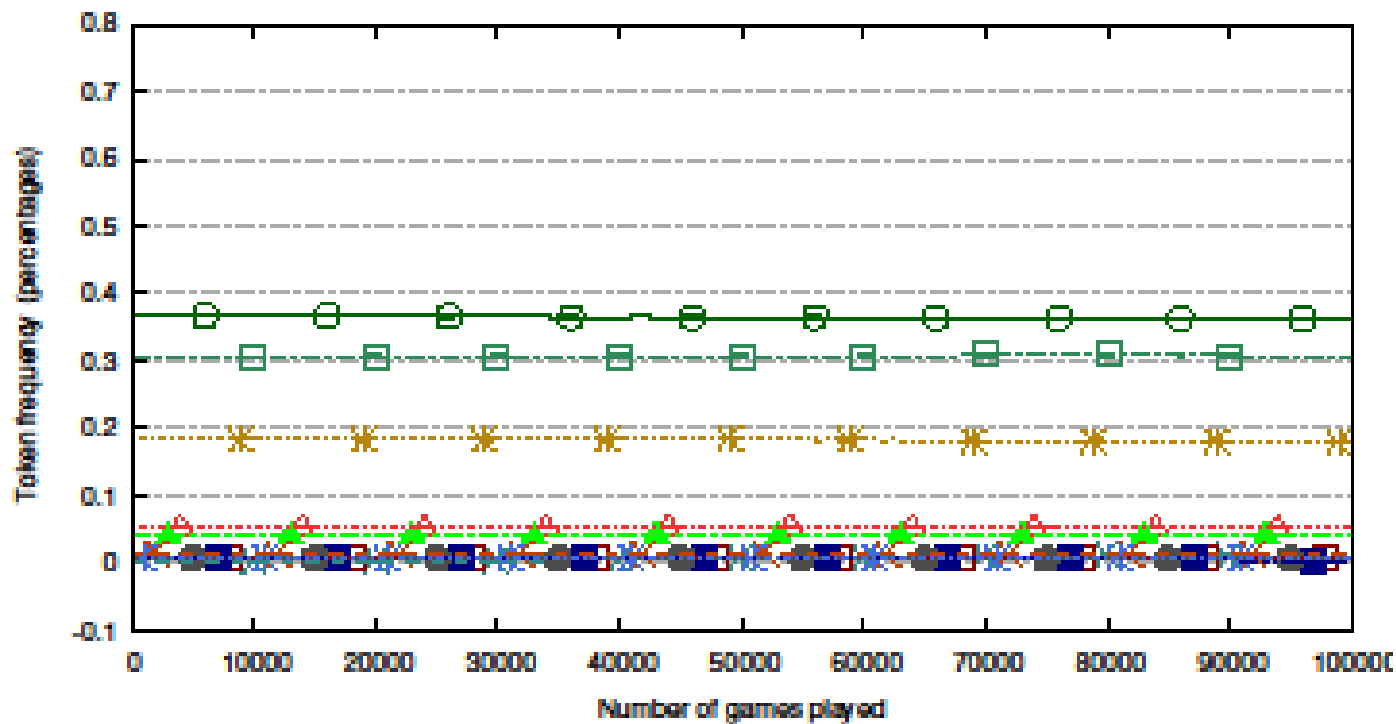
Results

- 8: token frequencies



Results

- 8: type frequencies

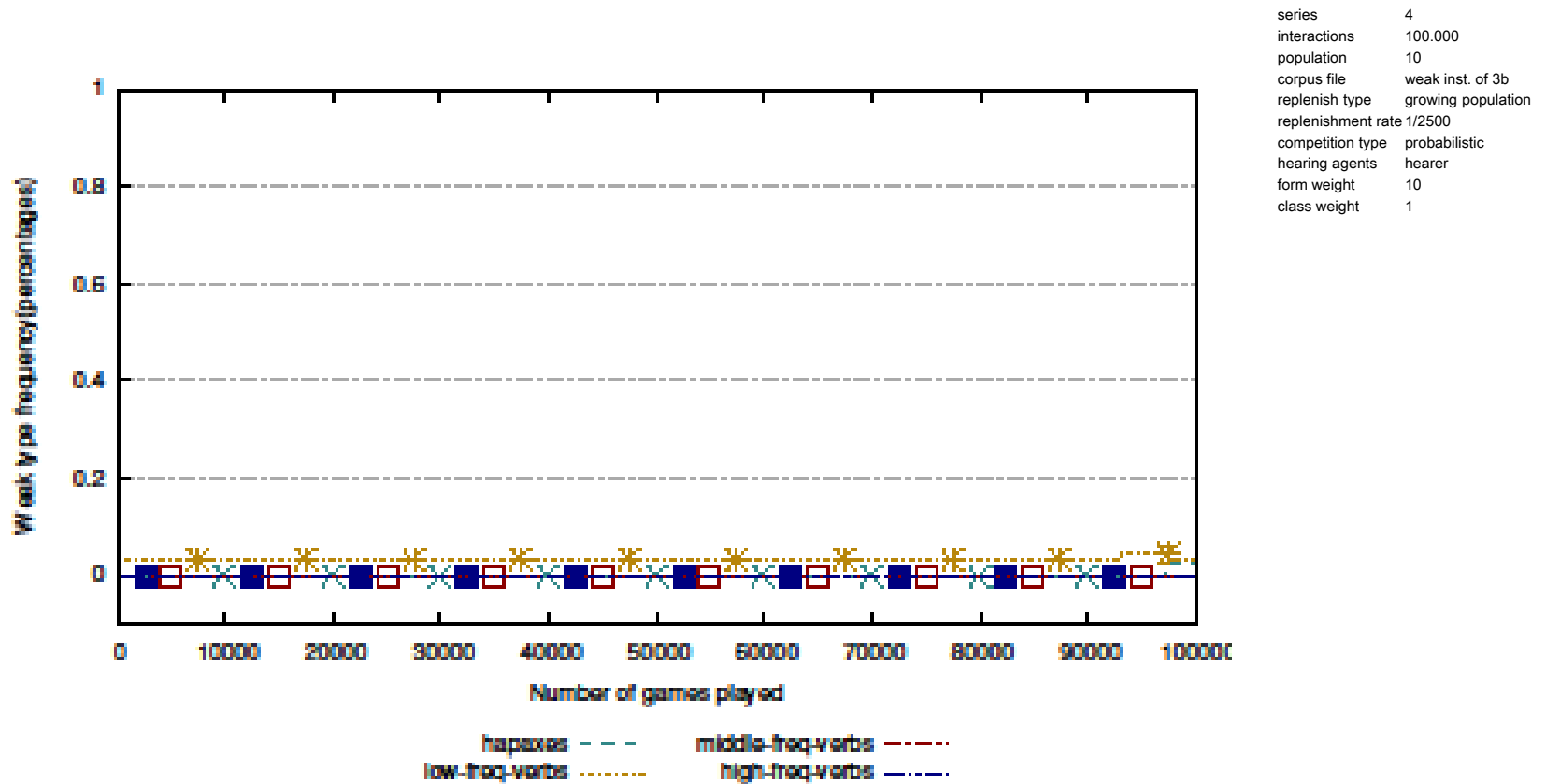


weak - - - 2a - - - 3a1 - - - 4a5a - - - 7a1 - - - 8a - - -
 1a ····· 2b - - - 3a2 - - - 6a - - - 7b ·····

series 4
 interactions 100.000
 population 10
 corpus file weak inst. of 3b
 replenish type growing population
 replenishment rate 1/2500
 competition type probabilistic
 hearing agents hearer
 form weight 10
 class weight 1

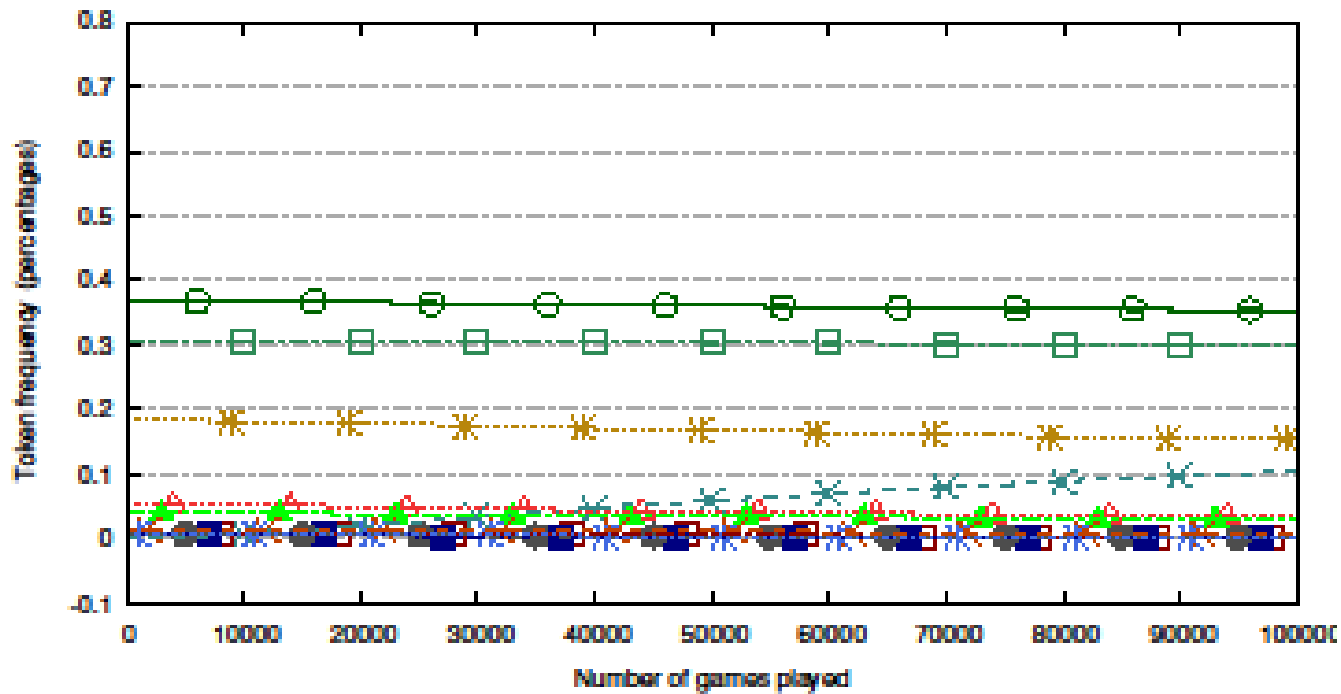
Results

- 8: frequency bins



Results

- 9: token frequencies



series	4
interactions	100.000
population	10
corpus file	weak inst. of 3b
replenish type	growing population
replenishment rate	1/250
competition type	probabilistic
hearing agents	hearer
form weight	10
class weight	1

weak - - - 2a - - - 3a1 - - - 4a5a - - - 7a1 - - - 8a - - -
 1a - - - 2b - - - 3a2 - - - 6a - - - 7b - - -

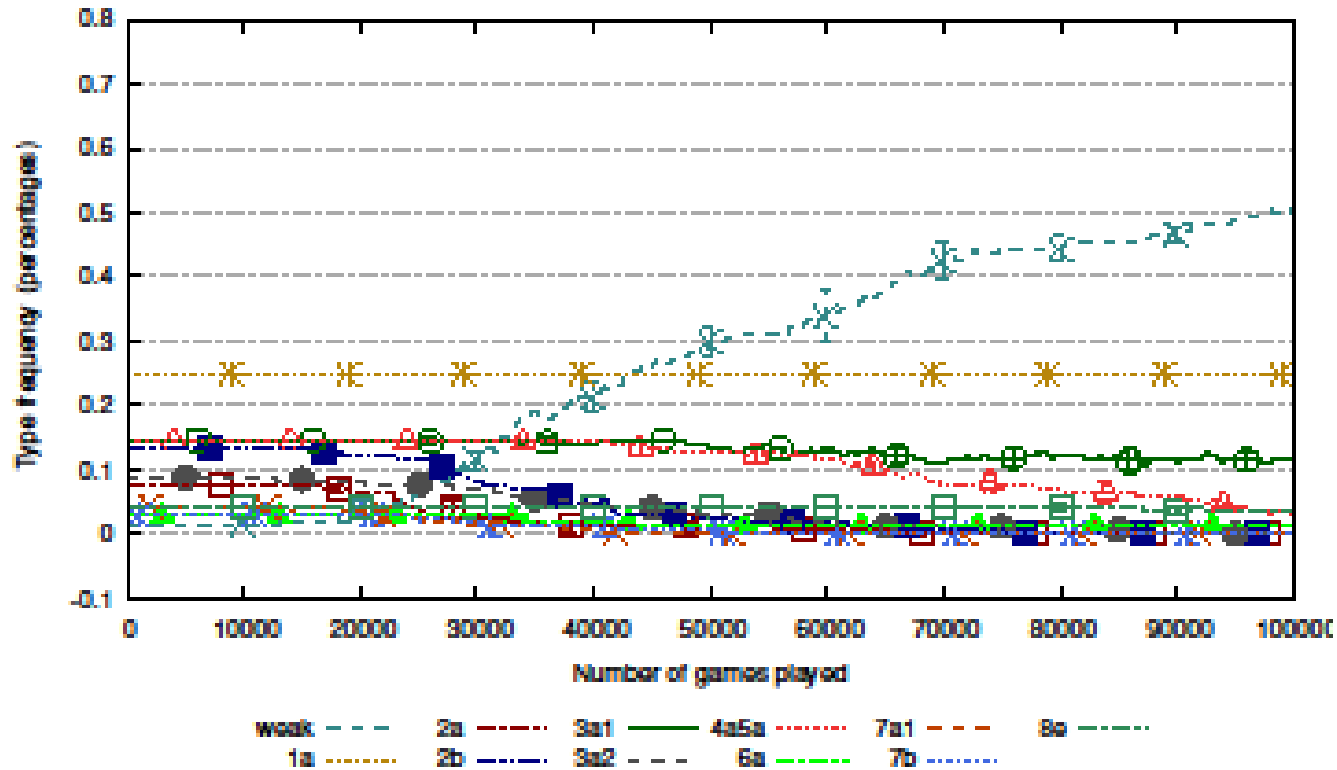


Results

- 9: type frequencies

```

series      4
interactions 100.000
population  10
corpus file  weak inst. of 3b
replenish type  growing population
replenishment rate 1/250
competition type  probabilistic
hearing agents  hearer
form weight  10
class weight  1
    
```



Results

- 9: frequency bins

series	4
interactions	100.000
population	10
corpus file	weak inst. of 3b
replenish type	growing population
replenishment rate	1/250
competition type	probabilistic
hearing agents	hearer
form weight	10
class weight	1

